Local Regularity Preserving Signal Denoising I: Hölder Exponents

Antoine Echelard, Jacques Lévy Véhel

Local Regularity Preserving Signal Denoising I: Hölder Exponents

A. Echelard and J. Lévy Véhel *

October 10, 2013

Abstract

We propose a denoising method that has the property of preserving local regularity, in the sense of local Hölder exponent. This approach is fitted to the processing of irregular signals, and gives specially relevant results for those displaying a local form of scale invariance known as localisability. A wavelet decomposition is used to measure and control the local Hölder exponent. The main ingredient of the algorithm is an estimator (which is of independent interest) of the time-dependent cut-off scale beyond which wavelet coefficients are mainly due to noise. Based on local regularity estimated from information below the cut-off scale, these small-scale coefficients -which govern the texture- are corrected so that the Hölder exponent of the denoised signal matches the one of the original signal. The processing is only slightly more complex than classical wavelet coefficients thresholding, resulting in fast computing times. Numerical experiments show the good performance of this scheme on various localisable signals.

1 Recalls on Hölder exponents and notations

We use the following notation throughout: f is a continuous-time signal that is always assumed to belong to the global Hölder space $C^{\epsilon}((0,1))$ for some

*A. Echelard and J. Lévy Véhel are with the Regularity Team, INRIA, Ecole Centrale Paris, Grande Voie des Vignes, 92290, Chatenay Malabry, France. e-mail: antoine.echelard@gmail.com, jacques.levy-vehel@inria.fr

$\epsilon > 0$. Recall that, when $\beta \in (0,1)$, $f \in C^\beta((0,1))$ means that there exists a constant C such that, for all $(x,y) \in (0,1)^2$, $|f(x) - f(y)| \leq C|x-y|^\beta$. More generally, when $m < \beta < m+1$ with m an integer, $f \in C^\beta((0,1))$ means that f is m times continuously differentiable and $|f^{(m)}(x) - f^{(m)}(y)| \leq C|x-y|^{\beta-m}$. The global Hölder exponent of f in the interval I, denoted $\alpha_f(I)$ or α_f, is the supremum of the β such that f belongs to $C^\beta(I)$. The local Hölder exponent of f at $x \in [0,1]$, denoted $\alpha_f(x)$ or $\alpha(x)$ is defined as $\alpha(x) = \lim_{\eta \to 0^+} \alpha_f(B(x,\eta))$, where $B(x,\eta)$ is the open ball centred at x with radius η. Thus, $\alpha(x)$ measures the regularity of f "around" x. A small value means an irregular behaviour, and *vice-versa*.
We assume without loss of generality that our signals are observed on $[0,1]$. When we write that f_n is an approximation at resolution n of f, we mean that f_n is a representation of f using 2^n samples. The letter h will always denote a non-decreasing function from $\mathbb{N}$ to $\mathbb{N}$ tending to infinity and such that $h(n) \leq n$ for all n. The abbreviation wlog means "without loss of generality", w.r.t. stands for "with respect to", iff is "if and only if", a.s. means "almost surely", i.i.d. abbreviates "independent and identically distributed", r.v. stands for "random variables", $\mathcal{N}(0,1)$ is used to denote the centred normal law with unit variance, and log is base-2 logarithm.

2 Statement of the problem

We consider the following situation: one observes $Y = X + B$ where X is the original signal and B is a white noise. One seeks an estimator $\hat{X}$ of X that has "good" properties. Obviously, one desirable property is that $\hat{X}$ is "close" to X in some sense. Typically, the error is measured by some risk function, and one wishes that, as the resolution n tends to infinity, this error tends to 0 at a fast rate. Additional properties are often useful. For instance, the celebrated method based on wavelet coefficients thresholding with the so-called universal threshold (see below for details) ensures that, with probability tending to one when n tends to infinity, $\hat{X}$ is at least as smooth as X. The significance of this feature is that, when presented with pure noise (*i.e.* when $Y = B$), the denoising scheme will indeed detect the absence of a signal (*i.e.* $\hat{X} = 0$).

The aim of the present work is to go beyond this property by designing a method that will ensure that, a.s., as n tends to infinity, $\hat{X}$ has the same local regularity -as measured by the local Hölder exponent- as X for a

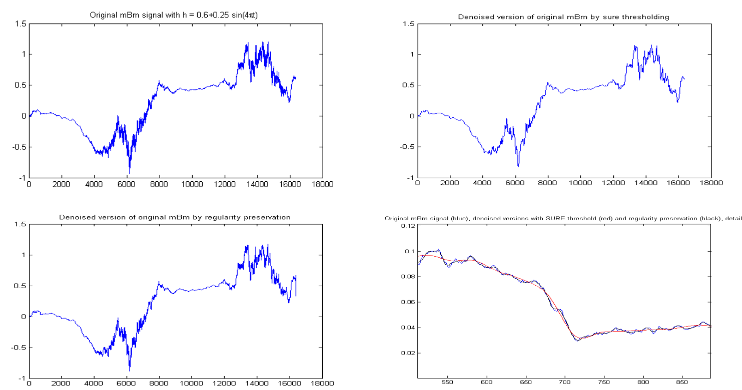

Figure 1: Original multifractional Brownian motion (top left), denoising with SURE thresholding (top right) and regularity preservation method (bottom left), and zoom on a superposition of the three signals (bottom right).

large class of (irregular) signals. There are several reasons for enforcing this constraint. First, it implies that, when presented with a "clean" signal (*i.e.* when $Y = X$), the denoising scheme will indeed yield $\hat{X} \simeq X$. This property is not shared by classical wavelet coefficients thresholding in the case where X is everywhere irregular (for instance, fractal). In this situation, $\hat{X}$ is significantly smoother than X (see Figure 1). Second, denoising is often only the first step in a chain of processings of the signal. While any decent scheme should guarantee that $\hat{X}$ and X are close, oversmoothing typically entailed by most methods may reduce the efficiency of the subsequent steps. This is the case when further processing is based on the study of irregularity. Examples include the analysis of biomedical signals (measuring the regularity of ECG allows one to assess the condition of the heart), financial records (where local regularity is related to the behaviour of agents and volatility of the market) or geophysical signals (e.g., for segmentation).

Of course, denoising everywhere irregular signals with the additional constraint of restoring the original regularity is more of a challenge, as it is difficult to distinguish the texture of the signal from the one of noise. We shall however see below that it is possible to ensure convergence of the estimated local Hölder exponents of $\hat{X}$ to the ones of X with good practical performance provided resolution is large enough.

The method that we develop in this work relies on a wavelet decomposition. We therefore briefly recall now some basic facts about wavelet-based denoising. This powerful approach has been very popular since the seminal papers [1, 2]. The essential idea is that, for many signals, only a few wavelet coefficients have significant magnitude, whereas the coefficients of B are uniformly distributed provided one uses an orthonormal wavelet basis. To denoise Y, it thus seems natural to replace its small coefficients by 0, and to keep or shrink large ones. This may be done in several ways, and there is a huge number of variants in this family of methods. In the sequel, we will denote ϕ and ψ the father and mother wavelets of a multiresolution analysis, and we assume that both functions are compactly supported and that ψ has sufficiently many vanishing moments. The wavelet coefficients of X are denoted $x = (x_{jk})_{j,k}$, those of Y, $y = (y_{jk})_{j,k}$, and $\hat{x} = (x_{jk})_{j,k}$ denotes the coefficients of the denoised signal $\hat{X}$. In the simplest case, thresholding is local, *i.e.* each coefficient is processed independently. The most well-known schemes are the hard- and soft-thresholding, where $\hat{x} = y\ 1_{|y| \geq \lambda}$ or $\hat{x} = \text{sign}(y)\ \max(0, |y| - \lambda)$. Popular choices for λ include the *minimax threshold* $\lambda^M = \hat{\sigma}\lambda_n^*$ where $\lambda_n^* = \inf_\lambda \sup_x \left\{ \frac{R_\lambda(x)}{2^{-n} + R_{\text{oracle}}(x)} \right\}$ and $\hat{\sigma}$ is the estimated standard deviation of B, $R_\lambda(x) = E((\hat{x_\lambda}(y) - x)^2$, and $R_{oracle}(x)$ is the ideal risk given by an oracle, such as DLP (diagonal linear projection) or DLS (diagonal linear shrinker) ones; *the universal threshold* $\lambda^U = \hat{\sigma} 2^{-n/2} \sqrt{2n}$, which ensures that, with probability tending to one when resolution tends to infinity, the zero signal contaminated with additive white Gaussian noise will be correctly estimated to zero; and the *SURE threshold* λ_j^S, obtained by considering the quantity

$$SURE(\lambda, X) = S - 2\sharp\{i, \|X_i\| \leq \lambda\} + \min(\|X_i\|, \lambda)^2$$

where $X_i = \frac{y_{j,k}}{\hat{\sigma}}$ and $S = 2^{j-1}$, and setting

$$\lambda_j^S = \underset{0 \leq \lambda \leq \lambda_j^U}{\text{argmin}} \left[SURE\left(\lambda, \frac{y_{j,k}}{\hat{\sigma}}\right) \right] \text{ with } \lambda_j^U = \hat{\sigma}\sqrt{2\log(2^j)}.$$

Other denoising rules include global thresholding, where all the coefficients of a given scale are processed in a single way, see, e.g. [3], and block thresholding [4]. The article [5] presents many other variants. They all typically perform well, although they tend to oversmooth the signal and also to introduce oscillations called "ringing effect". Ringing may be reduced significantly by various means. One is to use translation invariant wavelet coefficients

thresholding [6]. This however increases oversmoothing. A class of methods that refines thresholding using *a priori* information, which has been and still is the subject of substantial efforts, rely on Bayesian approaches. We do not go into details of these methods and refer instead the reader to [7, 8] and references therein.
To explain our concern in this work, we illustrate the oversmoothing effect of thresholding on a fractal signal in Figures 4 to 6. The original signal f is a Weierstrass function with exponent $\alpha = 0.5$, that was corrupted with additive Gaussian white noise to obtain the signal f_n (f and f_n are displayed in Figure 3), and denoised with hard-thresholding using a universal threshold (signal g_n) [1]. Each figure corresponds to a specific resolution. Although g_n is fairly close to f, it is clearly oversmoothed. As mentioned previously, this is a serious drawback in some applications where recovering the original regularity in addition to the overall shape is important. To explain in a heuristic way the phenomenon of oversmoothing in this particular case, recall the following result:

Proposition 1. *[9] Let $f \in C^{\epsilon}((0,1))$ with $\epsilon > 0$. Then*

$$1/2 + \alpha_f = \liminf_{j \to \infty} \min_{k \in \mathbb{Z}} \frac{\log |\langle f, \psi_{jk} \rangle|}{-j}, \tag{1}$$

where $\langle f, \psi_{jk} \rangle$ is the wavelet coefficient of f at scale j and location k. Thus, for the Weierstrass function, the coefficients at scale j are of the order of $2^{-j\alpha}$ or smaller. For large j, they are all are negligible w.r.t. the ones of the noise. As a consequence, the corresponding coefficients of f_n are essentially those of the noise and they get thresholded. This implies that g_n has vanishing coefficients at these scales and thus the original texture is lost. Our first aim will be to make this line of reasoning mathematically precise. This will allow us to explain another phenomenon observed on Figures 4 to 6: seen from “far away”, a signal denoised by wavelet thresholding typically looks smoother than the original (Fig. 4), but this impression diminishes as one zooms in (Fig. 6). A precise understanding of this feature will lead us to propose our new denoising scheme, which avoids this drawback.
Our approach fits in the paradigm proposed in [10]: rather than putting small coefficients to zero in the noisy signal, one tries to deduce their values from the ones of the large coefficients, which are assumed to be reliable. An

[1] Most other denoising based on thresholding yield the same kind of results.

example of implementation of this paradigm is total variation-based denoising, as exposed in [11]. In this variant, coefficients larger than a threshold are kept unchanged, while smaller ones are modified so as to minimize total variation. This permits to reduce ringing effects, but does not typically preserve texture. More generally, the formulation of [10] may be described as follows: let I be the set of indices for which the coefficients are larger than the threshold. Then the denoised signal $\tilde{f}_n$ is such that its coefficients with indices in I are not modified, whereas the other ones are chosen to be smaller in absolute value than a constant Q and such that an "energy" $\Phi(\tilde{f}_n)$ is minimum. While implementations of this paradigm improve on thresholding, in particular w.r.t. ringing effects, most are not satisfactory when it come to preserving regularity. In particular, they are not fitted to the processing of strongly textured signals as are fractal or, more generally, localisable signals. Recall that a process X is called localisable at u if there exists $\alpha > 0$ and a non-trivial process X'_u such that [12]

$$\lim_{r \to 0} \frac{X(u+rt) - X(u)}{r^\alpha} = X'_u(t). \tag{2}$$

The limit (2) may be taken either in finite dimensional distributions or distribution. Classical examples of localisable processes include multifractional Brownian motion, multifractional stable motion [13] and multistable motion [14]. Under general conditions, the local form X'_u is self-similar with stationary increments (sssi). Conversely, all sssi processes are localisable. Thus, localisable signals display a local form of scale invariance, and are typically everywhere irregular with a regularity that is time-dependent. They are often encountered in biomedicine, finance and geophysics. Local regularity is an important feature in such signals, as it bears crucial information on the state of the system.

Our denoising scheme follows a modified version of the paradigm of [10], with the difference that, rather than minimizing an energy functional, we seek to restore the regularity of the original signal, understood in a local sense and measured with the help of the Hölder exponent. This strategy is more relevant than thresholding for signals with non-sparse wavelet decomposition, as are localisable signals. For such a restoration to be possible, we need to be able to estimate the original regularity. This is performed by estimating, for each point, a cut-off scale beyond which the wavelet coefficients of Y are close to the ones of X. The Hölder exponent is then estimated from these wavelet coefficients. In contrast with the paradigm, we do not decide to keep coefficients unchanged if they are large enough, but rather when their scale is larger than the local cut-off. Coefficients below the cut-off scale are processed as follows: roughly speaking, a signal with exponent α at point t has wavelet coefficients above t that are smaller in absolute value than $2^{-j(\alpha+1/2)}$ at scale j. If a coefficient below the cut-off scale is larger than this value, it is thresholded, otherwise it is kept unchanged. This processing is based on interscale relations between the wavelet coefficients, and uses information on "known" coefficients to deduce the values of unknown ones. We note here that some works have already used regularity as a guide for denoising [15]. In contrast to our approach, they do not aim at recovering the regularity of the original signal. We also mention that interscale correlations of wavelet coefficients have already been exploited in a different way e.g. in [16]. Finally, [17] develops an approach that bears some similarities with ours in a different context.

The remainder of this article is organized as follows. In Section 3, we define a notion of "Hölder exponent in a range of scales" that is able to account for the perceived regularity of a signal at finite resolution. We examine in some details its properties in Section 4, in particular in relation with sampling and wavelet coefficients, and its links with Hölder exponents. Section 5 studies the behaviour of the Hölder exponent in a range of scales of a signal corrupted by Gaussian white noise, and Section 6 examines what happens in terms of regularity when a signal is denoised with hard-thresholding. The main theoretical result of this work is presented in Section 7: it provides an estimator of the location-dependent scale below which the wavelet coefficients of the original signal become negligible w.r.t. the ones of the noise. We believe that this result is of independent interest. With the help of this estimator, we present our denoising scheme in Section 8, and show that it is able to recover the regularity of the original signal. Finally, Section 9 displays experiments on localisable signals.

In a sequel to this paper, we extend the results obtained here when regularity is measured in a 2-microlocal sense rather than with Hölder exponents. 2-microlocal analysis gives a complete description of the local regularity of signals, and investigating denoising schemes in this frame provides further, sometimes unexpected, insights. For instance, we will prove that thresholding must typically introduce oscillations -in a well-defined mathematical sense- which are the source of the ringing effect. We will also show how to avoid this effect.

3 Exponents in a range of scales

Recall that, here as everywhere in the article, f is assumed to belong to $C^\epsilon((0,1))$ for some $\epsilon > 0$.

Case of a single function

In applications, one deals with signals sampled at finite resolution. As a consequence, Formula (1) cannot be applied directly to estimate α_f. This is a serious problem, as the value of the Hölder exponent is independent of an arbitrarily large but finite number of wavelet coefficients. Re-write (1) as follows:

$$\alpha_f = \liminf_{j \to \infty} \alpha_f(j),$$

where $\alpha_f(j)$ is defined as:

$$\alpha_f(j) = \min_{k \in \mathbb{Z}} \frac{\log |\langle f, \psi_{jk} \rangle|}{-j} - 1/2.$$

When the $\liminf$ above is a plain limit, it is not too difficult to estimate α_f. However, in most cases of interest, only a subsequence $\alpha_f(\sigma(j))$ tends to α_f. A further fact must be recognized, which is related to the visual significance of Hölder exponents: large scale wavelet coefficients do not influence the perceived smoothness. See Figure 2) for an illustration, where functions having same large scale (resp. small scales) coefficients are compared. This is just the obvious observation that large scale coefficients control the global aspect, whereas "texture" or roughness is governed by small scales ones. Similarly, the regularity of signals denoised by thresholding depends on the scale at which they are observed: the signal on Figure 4 looks significantly smoother than the original, the one on Figure 6 is more satisfactory; thus, looking at the denoised signal from far away may yield a satisfactory picture, whereas a close view reveals oversmoothing. In order to translates these facts into a mathematical framework, we introduce the notion of "Hölder exponents in a range of scales". Recall that here and everywhere in the article h denotes a non-decreasing function from $\mathbb{N}$ to $\mathbb{N}$ which tends to infinity and such that $h(n) \leq n$ for all n. The "Hölder exponent of f between scales $h(n)$ and n" is defined as:

$$\alpha_f(h(n), n) = \min_{j \in h(n) \ldots n, k \in \mathbb{Z}} \frac{\log |\langle f, \psi_{jk} \rangle|}{-j} - 1/2. \quad (3)$$

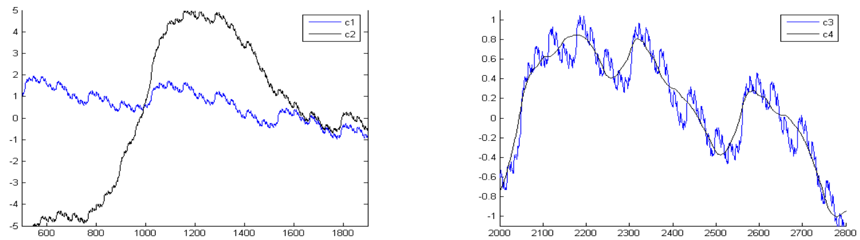

Figure 2: Perceived roughness depends on the amplitude of the wavelet coefficients at small scales. Left: both curves have same small scales coefficients, but differing ones in large scales. They produce the same impression of roughness. Right: both curves share the same coefficients at large scales, but differing ones at small scales. Their roughness appear to differ.

The indices $j \in \{h(n)...n\}$ are thus considered to be "texture scales", whereas the indices $j < h(n)$ are assumed to have no incidence on the perceived smoothness of the signal.

The following remarks are straightforward:

• $\liminf_{n\to\infty} \alpha_f(h(n), n) = \alpha_f$.

• If $(\alpha_f(h(n), n))_n$ converges, then its limit is α_f.

• $(\alpha_f(h(n), n))_n$ converges iff there exists a sequence $(\alpha_f(\sigma(k)))_k$ that tends to α_f and such that, for all n, there exists k with $\sigma(k) \in \{h(n)...n\}$.

• $(\alpha_f(h(n), n))_n$ converges iff there exists a sequence $(\alpha_f(\sigma(k)))_k$ which tends to α_f such that, for all k, $\sigma(k) \geq h(\sigma(k+1) - 1)$.

• If $h_1 \leq h_2$, then convergence of $(\alpha_f(h_2(n), n))_n$ implies convergence of $(\alpha_f(h_1(n), n))_n$.

• $\forall f,\ \exists\, h$ such that $(\alpha_f(h(n), n))_n$ converges.

• $\forall h,\ \exists\, f$ such that $(\alpha_f(h(n), n))_n$ diverges.

The last three points mean the following: the function h has to tend to infinity, but it may do so arbitrarily slowly. For any given f, it is always possible to choose h that tends to infinity sufficiently slowly so that $(\alpha_f(h(n), n))_n$ converges, but no single function h is sufficiently slow to fit all f.

Case of a sequence of functions

In practice, one does not deal with a single function, but with a sequence $(f_n)_n$, where each f_n is the approximation at resolution n of an underlying continuous-time signal f. Reasoning as above, the perceived roughness of each f_n will be controlled by the amplitude of the coefficients $\langle f_n, \psi_{jk} \rangle$ be-

tween some scale $h(n)$ and n. Then, if the sequence $(\alpha_{f_n}(h(n), n))_n$ tends to a limit l, one may expect that, for n sufficiently large, the perceived roughness will be comparable to the one of a function with exponent l . The situation here is however more complex than in the case of a single function. Indeed,
(a) a function h_0 such that $\alpha_{f_n}(h_0(n), n))_n$ converges does not always exist. For instance, define g_1 by $\langle g_1, \psi_{jk} \rangle = 2^{-j}$ and let $g_2 \equiv 0$. Set $f_{2n} = g_1$ and $f_{2n+1} = g_2$. It is easily seen that, for all h, $(\alpha_{f_{2n}}(h(n), n,))_n$ equals 0.5, whereas the sequence $(\alpha_{f_{2n+1}}(h(n), n))_n$ is identically infinite. As a consequence, the sequence $(\alpha_{f_n}(h(n), n))_n$ does not converge. Such an extreme behaviour is however rather rare,
(b) convergence of the sequence $\alpha_{f_n}(h_0(n), n))_n$ does not imply the one of $\alpha_{f_n}(h(n), n))_n$ when $h \leq h_0$. Furthermore, sequences $\alpha_{f_n}(h(n), n))_n$ may admit different limits depending on the choice of the sequences $h(n)$. This has practical implications, as we shall see in Section 6: let f_n denote the signal obtained by denoising using classical wavelet coefficients shrinkage with universal threshold at resolution n a signal f contaminated with additive Gaussian white noise. Then $\alpha_{f_n}(h(n), n))_n$ tends to α_f when h increases sufficiently slowly (which amounts to looking at the signal from "far away"). However, $\alpha_{f_n}(h(n), n)_n$ tends to $+\infty$ whenever $h(n)$ tends to infinity sufficiently fast: looking closely at the signal yields an impression of oversmoothing,
(c) in general, the limit of $\alpha_{f_n}(h(n), n))_n$, when it exists, depends on the analysing wavelet.

4 Estimated regularity of sampled signals

We assume from now on that each f_n is a sampling at resolution n of an underlying continuous-time signal f. If we accept (3) as valid definition of roughness in a range of scales, we need to relate $\alpha_{f_n}(h(n), n)$, $\alpha_f(h(n), n,)$ and α_f. Indeed, one can only compute coefficients $\langle f_n, \psi_{jk} \rangle$, which are just approximations of $\langle f, \psi_{jk} \rangle$ and one needs to examine how these approximations impact measured regularity. This further depends on how sampling is performed. We show that, provided that h increases slowly enough, the difference between the $\langle f_n, \psi_{jk} \rangle$ and the $\langle f, \psi_{jk} \rangle$ is sufficiently small so that $\alpha_f(h(n), n)$ is indeed well approximated by $\alpha_{f_n}(h(n), n)$, in two typical situations of sampling.

4.1 Impulse sampling

We first study the case where the samples are the values $f(k2^{-n})$, and consider two possibilities for defining the f_n.
Stepwise constant approximation: one possibility (used e.g. by the `cwt` function in the Wavelet Toolbox of Matlab) is to set:

$$f_n(t) = \sum_{i \in \mathbb{Z}} f(i2^{-n}) 1_{t \in [i2^{-n},(i+1)2^{-n})}. \tag{4}$$

In this case, one has

$$\langle f_n, \psi_{jk} \rangle = \sum_{i=0}^{2^n-1} f_n(i2^{-n}) \int_{2^{-ni}}^{2^{-n(i+1)}} \psi_{jk}(t) dt.$$

The following result, whose easy proof is omitted, allows one to estimate the error on the wavelet coefficients.

Proposition 2. *Let f_n be defined by* (4)*. Then there exits a constant C such that*

$$|\langle f_n, \psi_{jk} \rangle - \langle f, \psi_{jk} \rangle| \leq C \ 2^{-n\epsilon - j/2}.$$

Wavelet crime: wavelet coefficients are usually computed with the help of the fast wavelet transform [18]. One approximates $f(i2^{-n})$ by $\left\langle f, 2^{n/2}\phi_{ni} \right\rangle$ and sets

$$f_n = \sum_{i \in \mathbb{Z}} f(i2^{-n}) 2^{n/2} \phi_{ni}. \tag{5}$$

This is the so-called "wavelet crime" [19,20]. The easy proof of the following result is omitted:

Proposition 3. *There exists $C > 0$ such that for all $i \in \mathbb{Z}$:*

$$|\langle f_n, \phi_{ni} \rangle - \langle f, \phi_{ni} \rangle| \leq C 2^{-n\epsilon - n/2}.$$

We use the above result to prove the next statement:

Proposition 4. *Let $(f_n)_n$ be defined by* (5)*. There exists $C > 0$ such that, for all $j \leq n$,*

$$|\langle f, \psi_{jk} \rangle - \langle f_n, \psi_{jk} \rangle| \leq C 2^{-n\epsilon - j/2}.$$

Proof. For $j = n$, $\langle f_n, \psi_{nk} \rangle = 0$, while, since $f \in C^\epsilon(\mathbb{R})$, there exists C such that $|\langle f, \psi_{nk} \rangle| \leq C 2^{-n\epsilon - n/2}$.

For $j \leq n-1$, $\psi_{jk} \in \mathrm{Vect}\{\phi_{ni} : i \in \mathbb{Z}\}$ and thus:

$$\langle f_n, \psi_{jk} \rangle - \langle f, \psi_{jk} \rangle = \sum_{i \in \mathbb{Z}} \langle \langle f_n - f, \phi_{ni} \rangle \, \phi_{ni}, \psi_{jk} \rangle \, .$$

Using Proposition 3, one gets

$$|\langle f, \psi_{jk} \rangle - \langle f_n, \psi_{jk} \rangle| \leq C 2^{-n/2 - n\epsilon} \sum_{i \in \mathbb{Z}} |\langle \phi_{ni}, \psi_{jk} \rangle|$$

$$\leq C 2^{-n/2 - n\epsilon} \sum_{i \in \mathbb{Z}} 2^{n/2} \int_{-\infty}^{+\infty} |\phi(2^n t - i) \psi_{jk}(t)| \, dt.$$

Since ϕ has compact support, there exists M such that $t \to \sum_{i \in \mathbb{Z}} |\phi(2^n t - i)| < M$. Thus:

$$\begin{aligned} |\langle f, \psi_{jk} \rangle - \langle f_n, \psi_{jk} \rangle| &\leq C 2^{-n\epsilon} M \int_{-\infty}^{+\infty} |\psi_{jk}(t)| dt \\ &\leq C 2^{-j/2 - n\epsilon} M \int_{-\infty}^{+\infty} |\psi(t)| dt. \end{aligned}$$

□

Approximate wavelet coefficients and Hölder regularity: the results above allow one to give condition on the function h so that the Hölder exponents in a range of scales estimated on a sequence $(f_n)_n$ tend to the Hölder exponent of f.

Theorem 1. *Let f_n be the approximation of f at resolution n using* (5) *or* (4). *Assume that $h(n) \leq rn$ for an $r < 1$. Then,*

$$\alpha_f = \liminf_{n \to \infty} \alpha_{f_n}(h(n), n) \tag{6}$$

Proof. Since $f \in C^\alpha((0,1))$ for any $\alpha < \alpha_f$, Proposition 2 implies that $|\langle f_n, \psi_{jk} \rangle - \langle f, \psi_{jk} \rangle| \leq C_1 2^{-n\alpha - j/2}$. In addition, $|\langle f, \psi_{jk} \rangle| \leq C_2 2^{-j(\alpha + 1/2)}$. As a consequence,

$$|\langle f_n, \psi_{jk} \rangle| \leq (C_1 + C_2) 2^{-j(\alpha+1/2)},$$

$$\text{and } 1/2 + \alpha \le \liminf_{n\to\infty} \min_{j\in h(n)\ldots n, k\in\mathbb{Z}} \frac{\log|\langle f_n, \psi_{jk}\rangle|}{-j}.$$

This is true for all $\alpha < \alpha_f$, and we have proved one inequality. Choose now η small enough so that $r < \frac{\alpha_f - \eta}{\alpha_f + \eta}$. Let (j_l, k_l) be a sequence such that j_l tends to infinity and

$$|\langle f, \psi_{j_l,k_l}\rangle| \ge 2\ 2^{-j_l(\alpha_f+\eta+1/2)}.$$

Set $n_l = \lceil j_l \frac{\alpha_f+\eta}{\alpha_f-\eta} \rceil$. Since $h(n_l) \le rn_l$, there exists L such that for all $l \ge L$, $h(n_l) \le j_l \le n_l$, and thus

$$1/2 + \alpha_f(h(n_l), n_l) \le \frac{\log|\langle f_{n_l}, \psi_{j_l k_l}\rangle|}{-j_l}.$$

This implies that $|\langle f_{n_l}, \psi_{j_l k_l}\rangle| \ge 2^{-j_l(\alpha_f+\eta+1/2)}$. However, $f \in C^{\alpha_f-\eta/2}((0,1))$, and one may obtain a lower bound on $|\langle f_{n_l} - f, \psi_{j_l k_l}\rangle|$ with the help of Proposition 2:

$$|\langle f_{n_l}, \psi_{j_l,k_l}\rangle| \ge |\langle f, \psi_{j_l,k_l}\rangle| - C2^{-n_l(\alpha_f-\eta/2)-j_l/2}.$$

When $C2^{-n_l(\alpha_f-\eta/2)} \le \frac{1}{2}2^{-n_l(\alpha_f-\eta)} \le \frac{1}{2}2^{-j_l(\alpha_f+\eta)}$, one has $|\langle f_{n_l}, \psi_{j_l,k_l}\rangle| \ge \frac{1}{2}2^{-j_l(\alpha_f+\eta+1/2)}$. This means that $\alpha_f(h(n_l), n_l) \le \alpha_f + \eta$. The required inequality is obtained by taking the lim inf. $\square$

Theorem 1 implies that, if $h(n) \le rn$, and if $\alpha_{f_n}(h(n), n)$ converges, then its limit is α_f. In other words, the sequence $(\alpha_{f_n}(h(n), n))_n$ cannot converge to a "wrong limit". It however says nothing about the question whether the sequence converges. This is the topic of the next result. For a given f, we have seen that there always exists an h_0 such that the sequence $\alpha_f(h_0(n), n)$ converges. Theorem 2 shows that it is sufficient to choose h "slower" than h_0 to ensure that $(\alpha_{f_n}(h(n), n))_n$ will tend to α_f.

Theorem 2. *Let $h_0 : \mathbb{N} \to \mathbb{N}$ be a non-decreasing function tending to infinity with $h_0(n) \le n$ such that $\alpha_f = \lim_{n\to\infty} \alpha_f(h_0(n), n)$. Then, for any sequence u_n of integers such that $u_n \le rn$ with $r < 1$, and for any h satisfying the usual conditions and such that $h(n) \le h_0(u_n)$,*

$$\alpha_f = \lim_{n\to\infty} \alpha_{f_n}(h(n), n).$$

Proof. A lower bound on $\alpha_f(h(n), n)$ is obtained with the help of Proposition 2 as in the proof of Theorem 1:

$$\forall \eta > 0, \quad |\langle f_n, \psi_{jk}\rangle| \le (C_1 + C_2)2^{-j(\alpha_f-\eta+1/2)}$$

and thus $\alpha_f - 2\eta \leq \alpha_f(h(n), n)$ for large enough n. The upper bound also follows the same lines as in Theorem 1: for all $\eta > 0$, there exists a sequence (j_l, k_l) such that

$$|\langle f, \psi_{j_l,k_l}\rangle| \geq 2\ 2^{-j_l(\alpha_f+\eta+1/2)} \text{ and } j_l \geq h_0(j_{l+1}-1).$$

For $n \in \mathbb{N}$, consider $j_{l(n)}$ such that $u_n \in [j_{l(n)}, j_{l(n)+1}-1]$. Then $h(n) \leq h_0(u_n) \leq h_0(j_{l(n)+1}-1) \leq j_{l(n)} \leq u_n$, and thus $j_{l(n)} \in [h(n), u_n]$. As $u_n < n$, this implies that

$$1/2 + \alpha_f(h(n), n) \leq \frac{\log\left|\langle f_n, \psi_{j_{l(n)}k_{l(n)}}\rangle\right|}{-j_{l(n)}}.$$

Since $f \in C^{\alpha_f-\eta}(\mathbb{R})$, one may bound $\left|\langle f_n, \psi_{j_{l(n)}k_{l(n)}}\rangle\right|$ from below with Proposition 2. Using that $n \geq r^{-1}j_{l(n)}$, one gets

$$\left|\langle f_n, \psi_{j_{l(n)}k_{l(n)}}\rangle\right| \geq 2\ 2^{-j_{l(n)}(\alpha_f+\eta+1/2)} - C2^{-r^{-1}j_{l(n)}(\alpha_f-\eta+1/2)}.$$

Choose $\eta < \frac{1-r}{1+r}(\alpha_f + 1/2)$. Then $\alpha_f + \eta + 1/2 < r^{-1}(\alpha_f - \eta + 1/2)$ and, for large enough n,

$$\left|\langle f_n, \psi_{j_{l(n)}k_{l(n)}}\rangle\right| \geq 2^{-j_l(\alpha_f+\eta+1/2)}.$$

One concludes as before. $\square$

We have proved that it is always possible to choose h tending to infinity slowly enough so that $\alpha_f+1/2$ is the limit of both sequences $\left(\min_{j\in h(n)\ldots n, k\in\mathbb{Z}} \frac{\log\left|\langle f_n,\psi_{jk}\rangle\right|}{-j}\right)_n$ and $\left(\min_{j\in h(n)\ldots n, k\in\mathbb{Z}} \frac{\log\left|\langle f,\psi_{jk}\rangle\right|}{-j}\right)_n$. Furthermore, this is the only possible limit as soon as $h(n) < rn$ with $r < 1$.

4.2 Integral sampling

A more realistic modelling of the sampling of f is to consider that, rather than measuring the values $f(k2^{-n})$, one has access to mean values $\int f(t)c(2^n t - i)2^n dt$ where c is a positive function whose integral is equal to 1 that characterizes the sampling device [21]. All the results of Section 4.1 remain valid in this situation: indeed, by the mean-value theorem, for all n and i, there exists $t_n^i \in [i2^{-n}, (i+1)2^{-n}]$ such that $f(t_n^i) = \int f(t)c(2^n t - i)2^n dt$. Replacing $f(i2^{-n})$ by $f(t_n^i)$ in the proofs yields the result.

5 Estimating the regularity of a noisy signal

The aim of this section is to evaluate the Hölder regularity of a signal f which has been corrupted with an additive Gaussian white noise B with standard deviation σ_0. Since dividing a signal by a constant does not change regularity, one may assume wlog that $\sigma_0 = 1$. Let $g = f + B$ denote the observed noisy signal. We are not interested here in the theoretical Hölder exponent of g, but in the one estimated from a sampling between scales $h(n)$ and n. Note that since g is a distribution, and thus does not belong to $C^{\epsilon}(\mathbb{R})$ for any $\epsilon > 0$, the results of the previous section do not apply. Furthermore, the values of g at the points $k2^{-n}$ are not well defined. Following [21], we assume that the samples read:

$$g_n^i = \int f(t)c(2^n t - i)2^n dt + 2^{-n/2} b_n^i$$

where c is as in the previous section and the b_n^i are i.i.d. $\mathcal{N}(0,1)$ r.v. The approximate wavelet coefficients are

$$\langle g_n, \psi_{jk} \rangle = \langle f_n, \psi_{jk} \rangle + 2^{-n/2} b_{jk}^n, \tag{7}$$

with the b_{jk}^n i.i.d. $\mathcal{N}(0,1)$ r.v. The next lemmas will be useful.

Lemma 1. *Let g_n be defined by* (7). *For all $\epsilon > 0$, there exists almost surely $N \in \mathbb{N}$ such that, for all $n \geq N$ and all k,*

$$|\langle g_n, \psi_{nk} \rangle| \geq 2^{-n(\epsilon+1/2)}.$$

Lemma 2. *Let $\epsilon > 0$ and b_{jk}^n be i.i.d $\mathcal{N}(0,1)$ r.v. Almost surely, there exists $N \in \mathbb{N}$ such that, for all $n \geq N$,*

$$\max_{j \in [1..n], k \in [0...2^j]} |b_{jk}^n| \leq 2^{n\epsilon}.$$

The proof of the first lemma is straightforward, while the second one is well-known. The main result of this section is:

Theorem 3. *Let g_n be defined by* (7). *Then $\alpha_{g_n}(h(n), n)$ tends almost surely to 0 when n tends to infinity.*

Proof. Fix $\epsilon > 0$. Lemma 1 implies that, for n large enough, a.s., $\frac{\log|\langle g_n, \psi_{n0} \rangle|}{-n} \leq \epsilon + \frac{1}{2}$, and thus $\alpha_{g_n}(h(n), n) \leq \epsilon$. To prove the reverse inequality, we must show that: $\forall j, h(n) \leq j \leq n$ implies $|\langle g_n, \psi_{jk} \rangle| \leq 2^{-j(\frac{1}{2}-\epsilon)}$. By (7), $|\langle g_n, \psi_{jk} \rangle| \leq |\langle f_n, \psi_{jk} \rangle| + 2^{-n/2}|b_{jk}^n|$, and Lemma 2 implies the result. □

Applying Theorem 3 with $f = 0$ yields that the estimated Hölder exponent in a range of scales of sampled Gaussian white noise tends a.s. to 0 when n tends to infinity. It is well known that the Hölder exponent of a Gaussian white noise b is equal to -1/2. Thus, irrespective of h, the sequence $\alpha_{b_n}(h(n), n)$ does not tend to α_b. This does not contradict Theorem 2 as b does not belong to $C^\epsilon((0, 1))$. Also, Theorem 3 means that the estimated (as well as the theoretical) regularity of $f + B$ does not depend on the regularity of f.

6 Estimating the regularity of signals obtained through denoising by thresholding

In the previous section, we have studied the theoretical properties of regularity of a noisy signal. Here, we follow the same approach to obtain the regularity of signals denoised with hard thresholding. We consider the universal threshold and a level dependent one. Other thresholds or comparable schemes, such as the ones presented in Section 2, yield similar results.

Theorem 4. *Let g_n be defined by (7) and $\tilde{f}_n$ be the signal denoised by hard thresholding g_n with $\lambda_n = 2^{-n/2}\sqrt{2n \ln 2}$. Let $\sigma(n)$ be a non-decreasing integer sequence such that $\alpha_f = \lim_{n\to\infty} \alpha_f(\sigma(n))$. Assume that h verifies:*

$$\exists \epsilon_0 > 0, \forall n \in \mathbb{N}, \exists i \in \mathbb{N} : \sigma(i) \in \left[h(n), \frac{n}{1+2\alpha_f}(1-\epsilon_0)\right]. \tag{8}$$

Then $\alpha_{\tilde{f}_n}(h(n), n)$ tends in probability to α_f.

Assume that h verifies:

$$\exists \epsilon_0 > 0 : \forall n \in \mathbb{N},\ h(n) \geq \frac{n}{1+2\alpha_f}(1+\epsilon_0). \tag{9}$$

Then $\alpha_{g_n}(h(n), n)$ tends in probability to $+\infty$.

In the sequel, we will call the quantity $c_n := \frac{n}{1+2\alpha_f}$ the *cut-off scale*, even though this quantity is not necessarily an integer. The structure of the proof is as follows: the coefficients $\langle f, \psi_{jk}\rangle$ are smaller in absolute value than $C2^{-j(\alpha_f+1/2)}$. Those of the noise are of the order of $2^{-n/2}$. Thus, when $j > c_n$, the coefficients of f are buried in the noise. Lemma 3 makes this precise by showing that the coefficients corresponding to these values of j are thresholded. As a consequence, $\alpha_{\tilde{f}_n}(h(n), n)$ tends in probability to infinity

when (9) is verified. At scales j smaller than the cut-off scale, the amplitude of noise is typically small w.r.t. $2^{-j(\alpha_f+1/2)}$: Lemma 4 shows that coefficients in these scales remain not larger than $2^{-j(\alpha_f+1/2)}$. This implies the sequence $\alpha_{g_n}(h(n), n)$ remains larger than α_f. Finally, there exists coefficients of the order of $2^{-j(\alpha_f+1/2)}$. For such coefficients which lie between $h(n)$ and the cut-off scale, noise will again be small and will not modify their order of magnitude, as shown in Lemma 5. The intuitive meaning is that there are two cases where the perceived regularity of the denoised signal can be assessed:
- $h(n) < \sigma(i) < \frac{n(1-\epsilon_0)}{1+2\alpha_f}$: the signal is seen from far away. There exists values of the sequence σ between $h(n)$ and $\frac{n}{1+2\alpha_f}$. The corresponding coefficients of the denoised signal are of the same order of magnitude as the ones of the original signal, and $\alpha_{\tilde{f}_n}(h(n), n)$ remains close to α_f.
- $\frac{n(1+\epsilon_0)}{1+2\alpha_f} < h(n)$: the signal is seen from a close distance. With large probability, all the coefficients of the denoised signal at scales $j \geq h(n)$ vanish, and thus $\alpha_{\tilde{f}_n}(h(n), n) = +\infty$; in other words, the signal is oversmoothed.
This is exactly what is observed in Figures 4 to 6. Note that Theorem 4 does not cover all cases: when $h(n)$ is smaller than c_n, but sufficiently close to it so that no terms in the sequence σ belong to $[h(n), c_n]$, knowledge of α_f is insufficient to predict the behaviour of $\alpha_{\tilde{f}_n}(h(n), n)$. In "nice cases", e.g. when $\sigma(i) = O(i)$, this happens only when $h(n)$ is of the order of $\frac{n}{1+2\alpha_f}$ for an infinity of indices n.
We will need the following fact, a slight generalization of a classical result given, e.g., in [22], whose proof is omitted:

Fact 1. *Let $(\epsilon_n)_n$ be a positive sequence such that $\epsilon_n = o(n)$. Let $(z_n)_{n\in\mathbb{N}}$ be i.i.d. $\mathcal{N}(0,1)$ r.v. Then*

$$P\left(\max_{i\in[1\ldots 2^n]} |z_i| \leq \sqrt{2n\ln 2 - \ln \epsilon_n}\right) \to 1.$$

Lemma 3. *For any $\epsilon > 0$,*

$$P\left(\forall j \geq \frac{n}{1+2\alpha_f}(1+\epsilon) \ : \ \langle \tilde{f}_n, \psi_{jk}\rangle = 0\right) \to 1.$$

Proof. Choose η small enough so that $(1+\epsilon)\frac{\alpha_f+1/2-\eta}{\alpha_f+1/2} > 1$. Then, for n large enough, say $n \geq N_1$,

$$\frac{n}{1+2\alpha_f}(1+\epsilon)\left(\alpha_f + 1/2 - \eta\right) \geq \frac{n + \ln(8n\ln 2)}{2}.$$

Theorem 1 implies that there exists N_2 such that, for all $n \geq N_2$, all $j \geq \frac{n}{1+2\alpha_f}(1+\epsilon)$ and all k,

$$
\begin{aligned}
|\langle f_n, \psi_{jk}\rangle| &\leq 2^{-j(\alpha_f+1/2-\eta)} \\
&\leq 2^{-\frac{n}{1+2\alpha_f}(1+\epsilon)(\alpha_f+1/2-\eta)}.
\end{aligned}
$$

Thus, for $n \geq \max(N_1, N_2)$, $j \geq \frac{n}{1+2\alpha_f}(1+\epsilon)$ and all k,

$$
\begin{aligned}
|\langle f_n, \psi_{jk}\rangle| \leq 2^{-\frac{n+\log(8n\ln 2)}{2}} &\leq \frac{2^{-n/2}}{2\sqrt{2n\ln 2}} \\
&\leq 2^{-n/2}\sqrt{2n\ln 2}\left(1-\left(1-\frac{1}{2n\log 2}\right)^{1/2}\right). \\
&\leq \lambda_n - 2^{-n/2}\sqrt{2n\ln 2 - 1}.
\end{aligned}
$$

Using Fact 1 with $\epsilon_n = e$, one gets

$$
P\left(\forall j \in [1..n], k \in [1..2^j], |b_{jk}^n| \leq 2^{-n/2}\sqrt{2n\ln 2 - 1}\right) \to 1.
$$

This implies

$$
P\left(\forall j \geq \frac{n}{1+2\alpha_f}(1+\epsilon), k \in [1..2^j], |\langle g_n, \psi_{jk}\rangle| \leq \lambda_n\right) \to 1
$$

which is our result. □

Lemma 4. *For any $\eta > 0$, there exists $\epsilon > 0$ such that, for all sequence $h(n)$ tending to infinity, there exists a.s. $N \in \mathbb{N}$ such that for all $n \geq N$:*

$$
\forall j \in \left[h(n), \frac{n}{1+2\alpha_f}(1+\epsilon)\right], \left|\langle \tilde{f}_n, \psi_{jk}\rangle\right| \leq 2^{-j(\alpha_f+1/2-\eta)}.
$$

Proof. Fix $\eta > 0$. Choose ϵ such that

$$
\frac{1+\epsilon}{1+2\alpha_f}(\alpha_f + 1/2 - \eta) = 1/2 - \gamma,
$$

where $\gamma > 0$. Then, for all $j \leq \frac{(1+\epsilon)n}{1+2\alpha_f}$,

$$
2^{-j(\alpha_f+1/2-\eta)} \geq 2^{-n/2}\ 2^{\gamma n}.
$$

Lemma 2 implies that, a.s., there exists $N_1 \in \mathbb{N}$ such that

$$\forall n \geq N_1,\ 2^{-n/2}|b^n_{jk}| \leq \frac{1}{2}\ 2^{-j(\alpha_f+1/2-\eta)}.$$

Lemma 1 implies that there exists N_2 such that

$$\forall n \geq N_2,\ \forall j \geq h(n),\ |\langle f_n, \psi_{jk}\rangle| \leq \frac{1}{2}\ 2^{-j(\alpha_f+1/2-\eta)}.$$

Thus, by definition of g_n, for all $n \geq \max(N_1, N_2)$,

$$\forall j \in \left[h(n), \frac{n}{1+2\alpha_f}(1+\epsilon)\right], |\langle g_n, \psi_{jk}\rangle| \leq 2^{-j(\alpha_f+1/2-\eta)}. \tag{10}$$

One concludes by noting that $\left|\langle \tilde{f}_n, \psi_{jk}\rangle\right| \leq |\langle g_n, \psi_{jk}\rangle|$. $\square$

Lemma 5. *Assume* (8)*. Then, for all $\eta > 0$, a.s., there exists $N \in \mathbb{N}$ such that for all $n \geq N$, there exists $j_n \in \left[h_n, \frac{n}{1+2\alpha_f}(1-\epsilon_0)\right]$ and $k_n \in [0, 2^{j_n}]$ such that*

$$\left|\langle \tilde{f}_n, \psi_{j_n k_n}\rangle\right| \geq 2^{-j_n(1/2+\alpha_f+\eta)}.$$

Proof. Let (j_n, k_n) be a sequence such that $\alpha_f = \lim_{n\to\infty} \frac{\log\left|<f,\psi_{j_n k_n}>\right|}{j_n} + 1/2$ and $j_n \in \left[h_n, \frac{n}{1+2\alpha_f}(1-\epsilon_0)\right]$. Choose $\eta > 0$ small enough so that $(1-\epsilon_0)(1+\frac{\eta}{1/2+\alpha_f}) = 1-2\gamma$, where $\gamma > 0$. Then, for n large enough,

$$|\langle f, \psi_{j_n k_n}\rangle| \geq 3\ 2^{-j_n(\alpha_f+1/2+\eta)}\ \geq 3\ 2^{-n/2+\gamma n}.$$

By Proposition 4, there exists $C > 0$ such that, for all n,

$$|\langle f, \psi_{j_n k_n}\rangle - \langle f_n, \psi_{j_n k_n}\rangle| \leq C 2^{-n\epsilon_0/2 - j_n/2}.$$

For n large enough, $C2^{-n\epsilon_0/2} \leq 2^{\gamma n}$, and this implies that $|\langle f, \psi_{j_n k_n}\rangle - \langle f_n, \psi_{j_n k_n}\rangle| \leq 2^{-n/2+\gamma n}$. As a consequence,

$$|\langle f_n, \psi_{j_n k_n}\rangle| \geq 2\ 2^{-n/2+\gamma n}.$$

Lemma 2 then implies that, a.s., there exists $N \in \mathbb{N}$ such that, for all $n \geq N$, $2^{-n/2}\left|b^n_{j_n k_n}\right| \leq \frac{1}{2}|\langle f_n, \psi_{j_n k_n}\rangle|$, so that

$$|\langle g_n, \psi_{j_n k_n}\rangle| \geq 2^{-j_n(\alpha_f+1/2+\eta)},\ \text{a.s.}$$

For n large enough, $2^{-j_n(\alpha_f+\frac{1}{2}+\eta)} \geq 2^{-n/2+\gamma n} > \lambda_n$. Thus, a.s., the coefficients $\langle g_n, \psi_{j_n k_n}\rangle$ are not thresholded and $\left|\langle \tilde{f}_n, \psi_{j_n k_n} >\right| \geq 2^{-j_n(\alpha_f+1/2+\eta)}$. $\square$

Proof. of Theorem 4

• *Case where h verifies* (8)*:* fix $\eta > 0$. Lemma 4 implies that there exists ϵ such that

$$P\left(\forall j \in \left[h(n), \frac{n(1+\epsilon)}{1+2\alpha_f}\right],\ \frac{\log\left|\langle \tilde{f}_n, \psi_{jk}\rangle\right|}{-j} \geq \alpha_f + \frac{1}{2} - \eta\right) \to 1.$$

Lemma 3 yields

$$P\left(\forall j \in \left[\frac{n(1+\epsilon)}{1+2\alpha_f}, n\right],\ \frac{\log\left|\langle \tilde{f}_n, \psi_{jk}\rangle\right|}{-j} \geq \alpha_f + \frac{1}{2} - \eta\right) \to 1.$$

One deduces that

$$P\left(\alpha_{g_n}(h(n), n) \geq \alpha_f + 1/2 - \eta\right) \to 1.$$

Finally, Lemma 5 implies that

$$P\left(\alpha_{\tilde{f}_n}(h(n), n) \leq \alpha_f + 1/2 + \eta\right) \to 1.$$

• *Case where h verifies* (9)*:* Lemma 3 yields that

$$P\left(\forall j \geq \frac{n}{1+2\alpha_f}(1+\epsilon),\ < \tilde{f}_n, \psi_{jk} >= 0\right) \to 1.$$

As a consequence, $P(\forall j \in [h(n), n], < \tilde{f}_n, \psi_{jk} >= 0) \to 1$ and thus $P(\alpha_{\tilde{f}_n}(h(n), n) = +\infty) \to 1$. □

Figures 4 to 6 display the behaviour of denoised signals as the number of samples increases. They illustrate Lemmas 3 to 5 in the case of a particularly simple signal, namely the Weierstrass function, which possesses a global scaling behaviour. The original signal and the noisy one (which will be used in all experiments involving this function) are displayed on Figure 3. Beyond the cut-off scale $\frac{n}{1+2\alpha_f}$, shown as a dotted horizontal blue line on the figures, all coefficients are thresholded (Lemma 3). At smaller scales, noise is smaller than $2^{-j(\alpha_f+1/2)}$ and the coefficients remain smaller than $2^{-j(\alpha_f+1/2)}$ (Lemma 4). Furthermore, "large" coefficients, those of the order of $2^{-j(\alpha_f+1/2)}$, remain of the same order of magnitude (Lemma 5). When resolution tends to infinity, this sequence will converge to the original signal, but each denoised signal looks more regular than the original one.

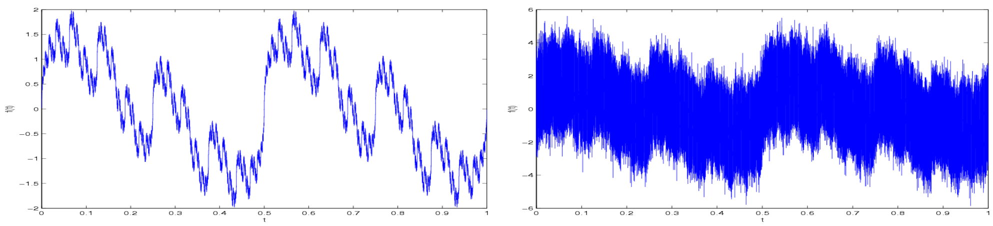

Figure 3: Original Weierstrass function sampled on 2^{19} points with exponent $\alpha_f = 0.5$ (left) and noisy version (right).

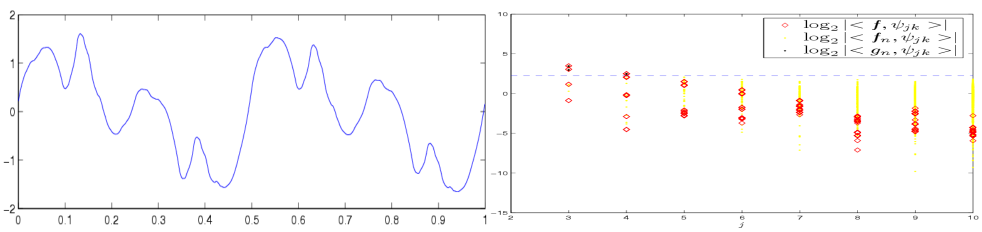

Figure 4: Denoised Weierstrass function sampled on 2^{11} points (left) and wavelet coefficients of the original, noisy and denoised versions (right)

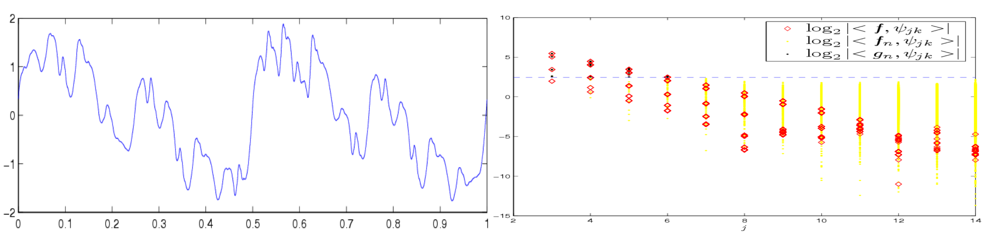

Figure 5: Denoised Weierstrass function sampled on 2^{15} points (left) and wavelet coefficients of the original, noisy and denoised versions (right)

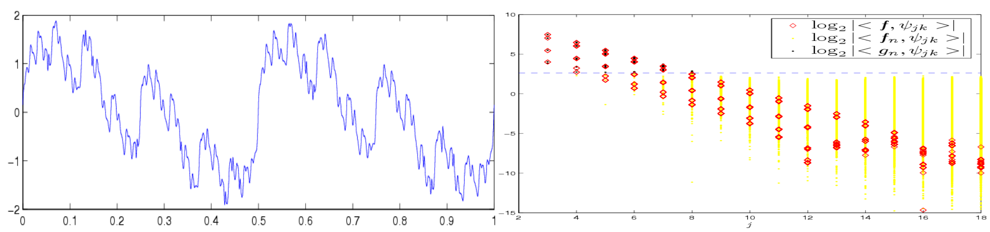

Figure 6: Denoised Weierstrass function sampled on 2^{19} points (left) and wavelet coefficients of the original, noisy and denoised versions (right)

The case of a level dependent threshold, where, at scale j, $\lambda_j = 2^{-j\delta}$, is treated in the following proposition, whose proof is similar to the ones of results above and is omitted.

Proposition 5. *Let g_n be defined by* (7) *and $\tilde{f}_n$ be the signal denoised with hard-thresholding and threshold $\lambda_j = 2^{-j\delta}$.*

1. *If $\delta > 1/2$ then $\alpha_{\tilde{f}_n}(h(n), n)$ tends a.s. to* 0.
2. *If $\delta < 1/2$ then $\alpha_{\tilde{f}_n}(h(n), n)$ tends a.s. to $+\infty$.*

In other words, a level-dependent threshold $2^{-j\delta}$ will either oversmooth the signal when $\delta < 1/2$ or yield a result with same estimated regularity as white noise when $\delta > 1/2$. One can show that a threshold of the form $2^{-j\delta - n/2}$ yields results analogous to the ones of the threshold $2^{-n/2}\sqrt{2n \ln 2}$.

7 Estimating the cut-off scale

In Section 8, we describe a denoising scheme that improves on thresholding in terms of regularity preservation. The feasibility of this scheme relies on the possibility of estimating the cut-off scale from the noisy signal. This section presents a way to do so in Corollary 6. This corollary follows Theorem 6, the main theoretical contribution of the present work. We will need following result (see, *e.g.*, [23], Theorem 2.7, p. 55).

Theorem 5. *Let $X_1, \ldots, X_n$ be independent r.v. Assume there exists positive constants $g_1, \ldots, g_n$, and T such that:*

$$\forall t \in [-T, T], \quad \mathbb{E}(e^{tX_k}) \leq e^{g_k \frac{t^2}{2}}, \quad k = 1 \ldots n.$$

Then, with $S_n = \sum_{k=1}^{n} X_k$ *and* $G_n = \sum_{k=1}^{n} g_k$,

$$\begin{array}{llll} P(S_n \geq x) & \leq & e^{-\frac{x^2}{2G_n}} & \text{for} \quad 0 \leq x \leq G_n T, \\ P(S_n \geq x) & \leq & e^{-\frac{Tx}{2}} & \text{for} \quad x \geq G_n T, \\ P(S_n \leq -x) & \leq & e^{-\frac{x^2}{2G_n}} & \text{for} \quad 0 \leq x \leq G_n T, \\ P(S_n \leq -x) & \leq & e^{-\frac{Tx}{2}} & \text{for} \quad x \geq G_n T, \end{array}$$

The following general result may be of independent interest. Its proof is given in the appendix.

Theorem 6. *Let* $(x_i)_{i\in\mathbb{N}}, (\sigma_i)_{i\in\mathbb{N}}$ *be two real sequences, with* $\sigma_i > 0$ *for all* i. *Assume that:*

1. $\beta = \liminf_{i\to\infty} \frac{-\log|x_i|}{i} > 0$;
2. *there exists a decreasing sequence* (ε_n) *such that* $\varepsilon_n = o\left(\frac{1}{n}\right)$ *when* $n \to \infty$ *and* $\frac{-\log|x_i|}{i} \geq \beta - \varepsilon_i$ *for all* i;
3. $0 < \delta' := \liminf\limits_{i\to\infty} \frac{-\log\sigma_i}{i} \leq \delta := \limsup\limits_{i\to\infty} \frac{-\log\sigma_i}{i} < \beta$.

For $n \in \mathbb{N}^*$, *let* Y *denote the random vector* $(y_1, \ldots, y_n)$ *where the* $(y_i)_i$ *are independent and, where, for each* i, y_i *is a Gaussian r.v. with mean* x_i *and variance* σ_n^2. *Set*

$$\mathcal{L}_n(p) = \frac{1}{(n-p+1)^2} \sum_{i=p}^{n} y_i^2.$$

Denote $p^* = p^*(n)$ *an integer such that*

$$\mathcal{L}_n(p^*) = \min_{1\leq p\leq n-b\log(n)} \mathcal{L}_n(p)$$

where $b > 1$ *is a fixed real number. Let finally* $q(n) = \frac{-\log\sigma_n}{\beta - \frac{1}{n}}$. *Then, for all* $a > 1$, *almost surely, for all* n *large enough,*

$$p^*(n) \leq q(n) + a\log(n). \tag{11}$$

Furthermore, if the sequence $(x_i)_i$ *satisfies the condition:*

4. *there exists a sequence of positive integers* (θ_n) *such that, for all sufficiently large* n *and all* $\theta \geq \theta_n$,

$$\frac{1}{\theta} \sum_{i=q-\theta}^{q-1} x_i^2 > b\sigma_n^2 \frac{1 - \frac{\delta_*}{\beta}}{(1 - \frac{\delta^*}{\beta})^2},$$

where $\delta_* \in (0, \delta')$ *and* $\delta^* \in (\delta, \beta)$,

then, for all $a > 1$, *almost surely, for all* n *large enough,*

$$p^*(n) \geq q(n) - \max(a \log(n), \theta_n). \tag{12}$$

Remark 1. *Assumption (3) and the definition of* q *imply that there exist* $0 < \rho' \leq \rho < 1$ *such that* $\rho' n \leq q(n) \leq \rho n$ *for all sufficiently large* n. *One may take* $\rho' = \frac{\delta_*}{\beta}$, $\rho = \frac{\delta^*}{\beta}$. *This fact and the assumption* $\varepsilon_n = o\left(\frac{1}{n}\right)$ *imply that* $\varepsilon_q < \frac{1}{n}$ *for sufficiently large* n *(in fact,* $\varepsilon_n = o\left(\frac{1}{n}\right)$ *can be replaced by the less restrictive condition* $\varepsilon_n < \frac{\rho'}{n}$*).*

Remark 2. *No assumption other than positivity is made on the sequence* (θ_n). *In particular, it does not have to tend to infinity. In the case where* $x_i = 2^{-i\beta}$, *one can take* $\theta_n = 2$.

Remark 3. *Condition 4 may be awkward to verify in practice. In many cases of interest, it can be replaced by the stronger but simpler condition* $\sigma_n^2 = o(\frac{1}{\theta} \sum_{i=q-\theta}^{q-1} x_i^2)$.

The meaning of the assumptions is as follows: (1) and (2) state that the x_i are bounded by $C2^{-i\beta+1}$. (3) essentially says that the variance σ_n^2 of the noise added to each x_i tends to 0 at a rate not faster than x_n, *i.e.* there is "enough noise". Under these conditions, the "normed energy" statistics $\mathcal{L}_n(p)$ has a minimum not larger than the cut-off level $q(n)$ where noise becomes predominant w.r.t. the signal. (4) means that we can group the x_i below $q(n)$ in blocks of a certain size θ in such a way that the energy of the block dominates the noise: this is a way of ensuring that there is a sufficient number of x_i which are large w.r.t. noise, *i.e.*, there is "enough signal". Then the minimum of $\mathcal{L}_n(p)$ is equal to the cut-off level within a logarithmic correction. We apply Theorem 6 to the following situation: the $(x_i)_i$ are the wavelet coefficients "above" a given point of a function X. One observes $Y = X + B$, with wavelet coefficients $(y_i)_i$ where B is a centred

Gaussian white noise. The problem is to estimate the value of the local Hölder exponent. In this setting, one has $\sigma_n = 2^{-n/2}$. Without noise and if the $(x_i)_i$ were all of the order of $C2^{-i\beta}$, then a simple linear regression on their logarithms would yield an estimate of β. In the presence of noise, one would observe, in logarithmic coordinates, the sum of points along a line with slope $-\beta$ (the $\log|x_i|$) and points on a horizontal line with ordinate $-n/2$ (the noise). Again, estimating β would be easy: it would amount to finding the level i^* where the line with slope $-\beta$ falls below the horizontal line. In general, however, the x_i are all smaller than $C2^{-i\beta}$, and only a subsequence is of this order of magnitude. Perhaps surprisingly, Theorem 6 and Corollary 6 say that, even in this situation, it is possible to estimate the cut-off i^* by using the statistics $\mathcal{L}_n(p)$ which is minimum close to i^* provided there are enough large x_i, *i.e.* (4) holds.

Corollary 6. *Let X be a function in $C^{\epsilon}((0,1)), \epsilon > 0$. Denote $(x_i)_i$ the wavelet coefficients of X "above" $t \in (0,1)$. Assume that θ_n defined in Theorem 6 is not larger than $b\log(n)$ for some $b > 1$ and all sufficiently large n. Let $Y = X + B$, with B a centred Gaussian white noise with unit variance. Set:*

$$\hat{\beta} = \frac{n}{2p^*} + \frac{1}{n},$$

where p^* is defined in Theorem 6. Then the following inequality holds almost surely for all sufficiently large n:

$$|\hat{\beta} - \beta| \leq 2b\beta^2 \frac{\log(n)}{n}.$$

Proof. From $\sigma_n = 2^{-n/2}$, one gets $q = \frac{n}{2\beta - \frac{1}{n}}$, or $\beta = \frac{n}{2q} + \frac{1}{n}$. Besides, $\hat{\beta} = \frac{n}{2p^*} + \frac{1}{n}$. In addition, the assumptions imply that $p^* \in [q - b\log(n), q + b\log(n)]$ a.s. for n large enough. Thus,

$$\begin{aligned}
|\hat{\beta} - \beta| &= n\frac{q - p^*}{2p^* q} \\
&\leq n\frac{b\log(n)}{q^2 - bq\log(n)} = \frac{nb\log(n)}{2n^2} 4(\beta - \frac{1}{n})^2 \\
&\leq 2b\frac{\log(n)}{n}(\beta - \frac{1}{n})^2 \quad \leq \quad 2b\beta^2\frac{\log(n)}{n}.
\end{aligned}$$

□

8 Local regularity preserving denoising

We assume that we have at our disposal a sequence $(s_n)_n$ that tends to α_f either in probability or a.s. Such a sequence is for instance provided by $\hat{\beta}$ defined in Corollary 6. This allows in turn to estimate the cut-off scale $\frac{n}{1+2\alpha_f}$ which is instrumental for our method. The idea is to keep coefficients at scales larger than $\frac{n}{1+2\alpha_f}$ and to diminish the ones at smaller scales so that they remain not larger in absolute value than $2^{-j(\alpha_f+1/2)}$. This allows one to recover the original regularity. More precisely, denoising is performed as follows:

Proposition 7. *Let g_n be defined by (7). Let $(s_n)_n$ be a sequence that tend a.s. (resp. in probability) to α_f. Let $\tilde{c}(n) = \frac{n}{1+2s_n}$. Define the denoised version $\tilde{f}_n$ by*

$$\left\langle \tilde{f}_n, \psi_{jk} \right\rangle = \begin{cases} \langle g_n, \psi_{jk} \rangle & \text{if } j \leq \tilde{c}(n) \\ \min\left(\left| \langle g_n, \psi_{jk} \rangle \right|, 2^{-j(s_n+1/2)} \right) sgn(\langle g_n, \psi_{jk} \rangle)) & \text{if } j > \tilde{c}(n). \end{cases} \tag{13}$$

Then $\alpha_{\tilde{f}_n}(h(n), n)$ tends a.s. (resp. in probability) to α_f.

Proof. Fix $0 < \eta < \alpha_f$. Inequality (10) always holds: there exists $n \in \mathbb{N}$ and $\epsilon > 0$ such that a.s., for n large enough,

$$\forall j \in \left[h(n), \frac{n}{1+2\alpha_f}(1+\epsilon) \right], \left| \langle g_n, \psi_{jk} \rangle \right| \leq 2 \; 2^{-j(\alpha_f+1/2-\eta)}.$$

Since $\tilde{c}(n)$ is a.s. equivalent to $\frac{n}{1+2\alpha_f}$, $\tilde{c}(n) \leq \frac{n}{1+2\alpha_f}(1+\epsilon)$ for n large enough. As a consequence,

$$\forall j \in [h(n), c(n)], \left| \left\langle \tilde{f}_n, \psi_{jk} \right\rangle \right| \leq 2 \; 2^{-j(\alpha_f+1/2-\eta)}.$$

Furthermore, for $j > \tilde{c}(n)$, $\left| \left\langle \tilde{f}_n, \psi_{jk} \right\rangle \right| \leq 2^{-j(s_n+1/2)}$ and thus a.s. for n large enough,

$$\forall j > \tilde{c}(n), \left| \left\langle \tilde{f}_n, \psi_{jk} \right\rangle \right| \leq 2^{-j(\alpha_f+1/2-\eta)}.$$

We have thus obtained that, for all $j \in [h(n), n]$, $\left| \left\langle \tilde{f}_n, \psi_{jk} \right\rangle \right| \leq 2 \, 2^{-j(\alpha_f+1/2-\eta)}$. This implies that $\alpha_{\tilde{f}_n}(h(n), n) \geq \alpha_f - \eta$.

As shown in the proof of Theorem 3, there exits a.s. $N \in \mathbb{N}$ such that for all $n \geq N$, $\left| \langle g_n, \psi_{n0} \rangle \right| \geq 2^{-n(\eta+1/2)}$, which implies that $\left| \left\langle \tilde{f}_n, \psi_{n0} \right\rangle \right| =$

$2^{-j(s_n+1/2)}$. Since s_n tends a.s. to α_f, $\left|\left\langle \tilde{f}_n, \psi_{n0} \right\rangle\right|$ tends a.s. to $2^{-j(\alpha_f+1/2)}$ and thus $\alpha_{\tilde{f}_n}(h(n), n) \leq \alpha_f + \eta$ a.s. when $n \to \infty$.
The proof when $(s_n)_n$ converges in probability is similar. $\square$

The procedure above may introduce jumps between coefficients at scales $j \leq \tilde{c}(n)$ and $j > \tilde{c}(n)$. To avoid this, we replace $2^{-j(s_n+1/2)}$ by $2^{K_n - j(s_n+1/2)}$, where $(K_n)_n$ is a bounded random sequence. Before explaining how to choose $(K_n)_n$, we prove that this modification does not impact regularity.

Corollary 8. *With the same notations as in Proposition 7, let $(K_n)_n$ be a sequence such that there exists $A, B > 0$ verifying: a.s., there exists $N \in \mathbb{N}$ such that, for all $n \geq N$, $K_n \in [A, B]$ (resp. $P(A \leq K_n \leq B) \to 1$). Define $\tilde{f}_n$ by*

$$\left\langle \tilde{f}_n, \psi_{jk} \right\rangle = \begin{cases} \langle g_n, \psi_{jk} \rangle & \text{if } j \leq \tilde{c}(n) \\ \min\left(\left|\langle g_n, \psi_{jk} \rangle\right|, 2^{K_n - j(s_n+1/2)}\right) sgn(\langle g_n, \psi_{jk} \rangle)) & \text{if } j > \tilde{c}(n). \end{cases}$$

Then $\alpha_{\tilde{f}_n}(h(n), n)$ tends a.s. (resp. in probability) to α_f.

Proof. Let $\hat{f}_n$ denote the function defined by (13).
Almost sure situation: assume that $K_n \in [A, B]$ is verified. We apply Proposition 7 to $2^A f$ and $2^B f$. Noting that $\alpha_{2^A f} = \alpha_{2^B f} = \alpha_f$, one sees that the same sequence s_n may be used. Thus both $\alpha_{\widehat{2^A f_n}}(h(n), n)$ and $\alpha_{\widehat{2^B f_n}}(h(n), n)$ tends a.s. to α_f. The result then follows from the inequalities

$$\alpha_{\widehat{2^A f}_n}(h(n), n) \leq \alpha_{\tilde{f}}(h(n), n) \leq \alpha_{\widehat{2^B f}_n}(h(n), n).$$

Situation in probability: Applying Proposition 7 to $2^A f$ and $2^B f$, one gets that $\alpha_{\widehat{2^A f_n}}(h(n), n)$ and $\alpha_{\widehat{2^B f_n}}(h(n), n)$ tend in probability to α_f. Since $\alpha_{\widehat{2^A f}_n}(h(n), n) \leq \alpha_{\tilde{f}}(h(n), n) \leq \alpha_{\widehat{2^B f}_n}(h(n), n)$, the result follows. $\square$

Experiments suggest that taking K_n equal to the offset in the regression line of the logarithm of the absolute values of the wavelet coefficients w.r.t. scale is a reasonable choice.

9 Numerical Experiments

As said in the first section, our approach is specially fitted to the case of localisable functions, which are irregular signals that fulfil a weak form of local scale invariance. We present in this section some results of denoising on such signals, and compare them with classical wavelet coefficients thresholding. It is well-known that the minimax and universal threshold are not well adapted when the wavelet coefficients are not sparse enough [16], and that the SURE threshold is better fitted for signals with small scale details, as are localisable processes. In order to make fair comparisons, we thus use soft-thresholding using the SURE threshold for classical thresholding.

Let us first recall that, contrarily to classical thresholding, our scheme will not, in principle, modify a locally scaling signal which is not contaminated by white noise, since, in this case, the wavelet coefficients are aligned, and the cut-off scale will be the maximal one. In practice, because of estimation issues, some minor changes will occur, which are much less visible than what is produced by classical thresholding. See Figure 1.

We begin with a signal with global scale invariance, namely the Weierstrass function, in Figures 7 to 9. One sees that regularity is recovered after denoising except on Figure 7, where the resolution is too low for the cut-off scale to be estimated with sufficient precision. It is interesting to contrast these results with the ones in Figures 4 to 6).

The subsequent experiments are on localisable signals. These signals are random, and we always choose the added Gaussian white noise to be independent of the signal. Our first example is multifractional Brownian motion (mBm) [14]. This is an extension of well-known fractional Brownian motion where the Hurst exponent is allowed to vary with time. This process has become a popular model in finance [24], geophysics [25], internet traffic modelling [26]and biomedicine [27]. We consider an mBm with local Hölder exponent $\alpha(t) = 0.6 + 25\sin(4\pi t)$ sampled on 2^{14} points. We display the original and noisy versions on Figure 10, along with regularity preservation denoisings but where the cut-off levels are fixed to 7 and 11, and finally the denoisings with regularity preservation and SURE thresholding. This signal was also used in Figure 1.

Let us finally consider a multifractional multistable process (mfmsp). This is a localisable process whose local form at each time is a well-balanced linear fractional stable motion [13]. An mfmsp depends on two functional parameters: the first one, denoted γ, controls the intensity of jumps, and ranges in $(0,2)$ (a small γ means a larger intensity of jumps). The second one, α, ranges in $(0,1)$, and controls roughness. Figure 11 displays an mfmsp with $\alpha(t) = 0.5 + 0.3\sin(4\pi t)$ and $\gamma(t) = 0.9 + t$, a noised version, and denoisings using regularity preservation and SURE thresholding.

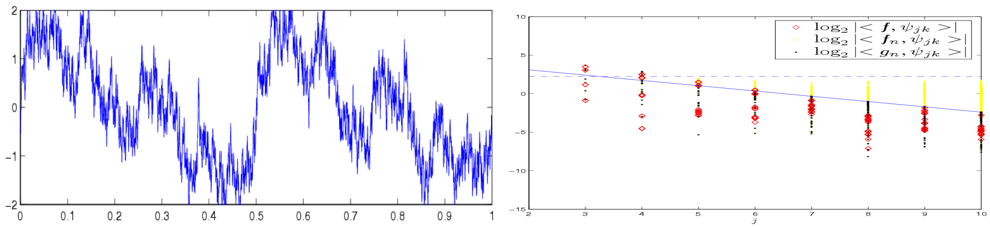

Figure 7: Denoised Weierstrass function sampled on 2^{11} points (left) and wavelet coefficients of original, noisy and denoised signals (right).

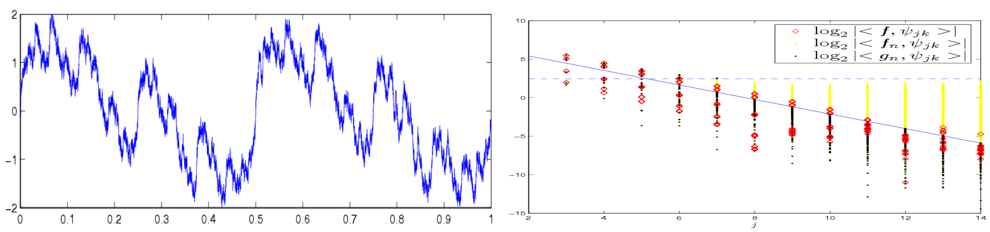

Figure 8: Denoised Weierstrass function sampled on 2^{15} points (left) and wavelet coefficients of original, noisy and denoised signals (right).

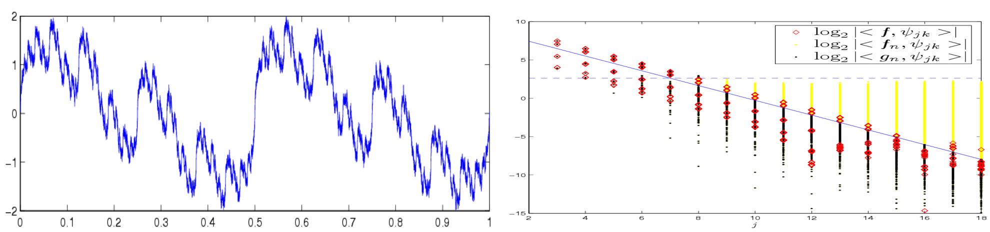

Figure 9: Denoised Weierstrass function sampled on 2^{19} points (left) and wavelet coefficients of original, noisy and denoised signals (right).

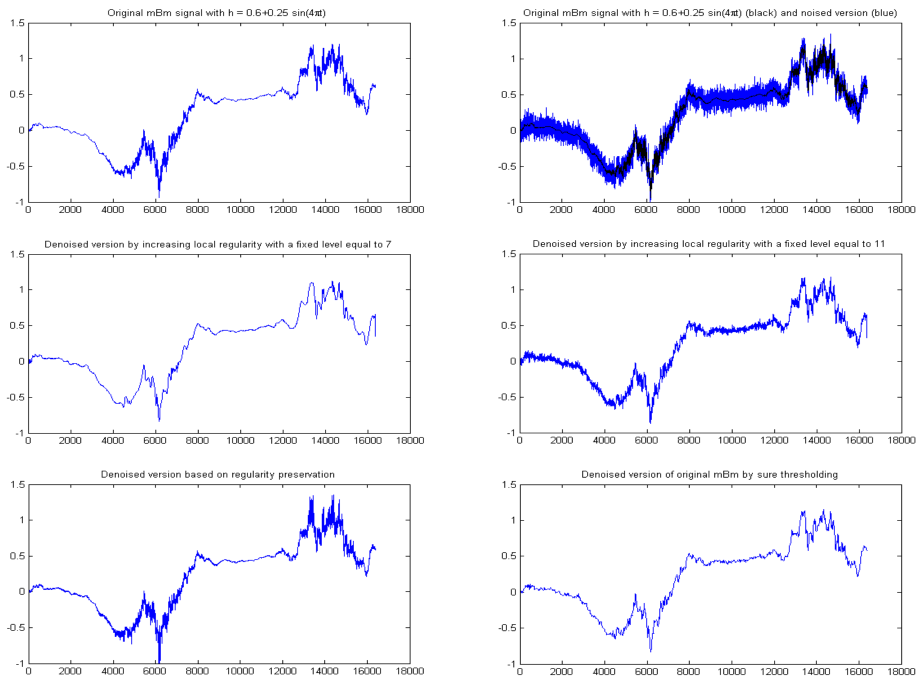

Figure 10: mBm sampled on 2^{14} points (top left), superimposed with noisy version (top right), denoising based on regularity preservation with fixed level equal to 7 (middle left) and 11 (middle right), denoised versions with regularity preservation (bottom left) and SURE thresholding (bottom right).

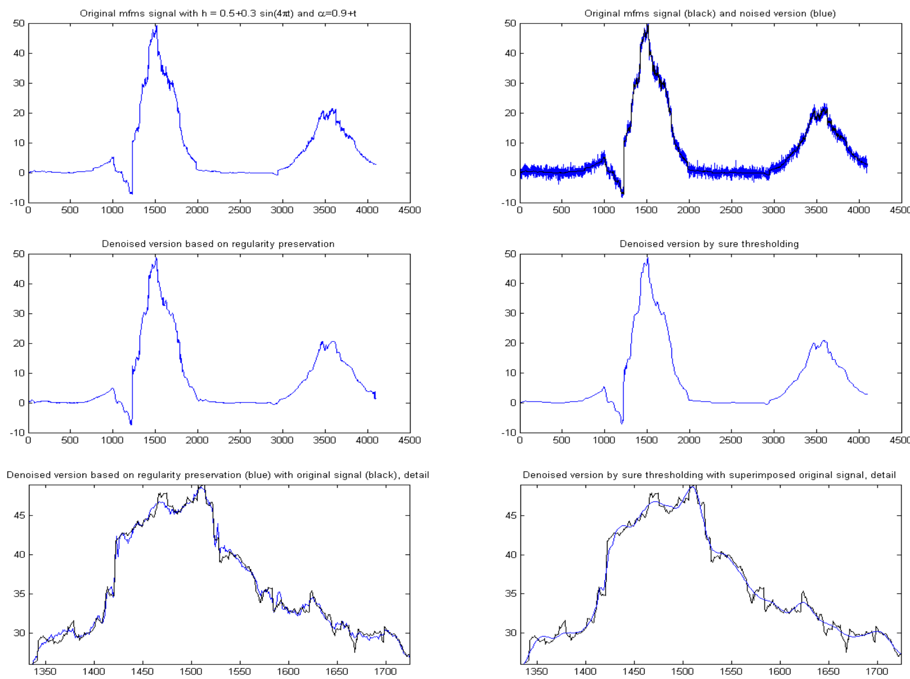

Figure 11: mfmsp sampled on 2^{12} points (top left), superimposed with noisy version (top right), denoised version based on regularity preservation (middle left) and SURE thresholding (middle right), zooms on denoised versions with regularity preservation (bottom left) and SURE thresholding (bottom right).

Appendix: proof of Theorem 6

We will need the following elementary lemma:

Lemma

Let $z = N(0,1)$ and $\Omega = z^2 + 2\lambda z - 1$, where $\lambda \in \mathbb{R}$.

Then

$$\mathbb{E}(e^{t\Omega}) \leq e^{\frac{gt^2}{2}} \tag{14}$$

for $t \in \left[-\frac{1}{4}, \frac{1}{4}\right]$ and $\begin{array}{rcl} g & = & 16\log 2 - 8 + 8\lambda^2 \ . \\ & =: & C + 8\lambda^2 \end{array}$

Proof. of the lemma. Simple computations yield $\mathbb{E}(e^{t\Omega}) = \frac{e^{t\left(\frac{\lambda^2}{1-2t} - (1+\lambda^2)\right)}}{\sqrt{1-2t}}$, from which (14) follows at once. □

Proof. of Theorem 6.

Set $\lambda_i = \frac{x_i}{\sigma_n}, \Omega_i = z_i^2 + 2\lambda_i z_i - 1$, where $z_i = \frac{y_i - x_i}{\sigma_n}$.

Proof of (11)

We want to show that, under assumptions 1) - 3), almost surely, $\mathcal{L}_n(q(n)) - \mathcal{L}_n(p(n)) \leq 0$ for large enough n, for any sequence $p(n)$ such that $q(n) + a\log(n) < p(n) \leq n - b\log(n)$ for any fixed $a > 1, b > 1$. Take such a sequence $p(n)$. Note that, in particular, $p(n) > q(n)$. For simplicity, we shall write $\mathcal{L}$, p, q in place of $\mathcal{L}_n$, $p(n)$, $q(n)$.

$\mathcal{L}(q) - \mathcal{L}(p) > 0$ is equivalent to:

$$X := (n-p+1)^2 \sum_{i=q}^{p-1} y_i^2 + (q-p)(2n-p-q+2) \sum_{i=p}^{n} y_i^2 > 0.$$

Define:

$$\begin{array}{rcll} X_i & = & \sigma_n^2 (n-p+1)^2 \Omega_i & \text{for} \quad i = q \ldots p-1, \\ X_i & = & \sigma_n^2 (q-p)(2n-p-q+2)\Omega_i & \text{for} \quad i = p \ldots n, \\ X_i & = & 0 & \text{for} \quad i = 0 \ldots q-1. \end{array}$$

On the one hand, one has:

$$
\begin{aligned}
S_n &:= \sum_{i=1}^{n} X_i \\
&= X-(n-p+1)^2\sum_{i=q}^{p-1} x_i^2 + (p-q)(2n-p-q+2)\sum_{i=p}^{n} x_i^2 \\
&\quad + \sigma_n^2(p-q)(n-p+1)(n-q+1) \\
&=: X+x
\end{aligned}
$$

Thus $X>0$ iff $S_n>x$.
On the other hand:

- for $i=q\dots p-1$:

  $$\mathbb{E}(e^{tX_i}) \le e^{g_i\frac{t^2}{2}} \text{ for } t\in\left[-\frac{1}{4\sigma_n^2(n-p+1)^2},\frac{1}{4\sigma_n^2(n-p+1)^2}\right] =: [-T_1,T_1]$$

  and $g_i = C\sigma_n^4(n-p+1)^4 + 8\sigma_n^2 x_i^2(n-p+1)^4$.

- for $i=p\dots n$:

  $\mathbb{E}(e^{tX_i}) \le e^{g_i\frac{t^2}{2}}$

  $$\text{for } t\in\left[\frac{1}{4\sigma_n^2(q-p)(2n-p-q+2)},-\frac{1}{4\sigma_n^2(q-p)(2n-p-q+2)}\right] =: [-T_2,T_2]$$

  and $g_i = C\sigma_n^4(p-q)^2(2n-p-q+2)^2 + 8\sigma_n^2 x_i^2(p-q)^2(2n-p-q+2)^2$.

We shall apply Theorem 5. In that view, we compute

$$
\begin{aligned}
G_n := \sum_{i=1}^{n} g_i &= C\sigma_n^4(p-q)(n-p+1)[(n-p+1)^3+(p-q)(2n-p-q+2)^2] \\
&+ 8\sigma_n^2[(n-p+1)^4\sum_{q}^{p-1} x_i^2 + (p-q)^2(2n-p-q+2)^2\sum_{i=p}^{n} x_i^2]
\end{aligned}
$$

We also need to compute the minimum of T_1 and T_2, so as to set $T = \min(T_1,T_2)$.

Now $T_2 < T_1$ iff $f(n,p,q) := (p-q)(2n-p-q+2) - (n-p+1)^2 > 0$. f is an increasing function of p, and one finds that:

$$\begin{aligned} T &= T_1 \quad \text{if} \quad n \geq p \geq \tfrac{\sqrt{2}}{2}q + \left(1 - \tfrac{\sqrt{2}}{2}\right)(n+1) =: \widetilde{p} \\ &= T_2 \quad \text{if} \quad q < p \leq \tfrac{\sqrt{2}}{2}q + \left(1 - \tfrac{\sqrt{2}}{2}\right)(n+1). \end{aligned}$$

The next step is to compare x and $G_n T$

- if $p \leq \widetilde{p}$:

$$G_n T - x = G_n T_2 - x =: A + B + C,$$

  where:

$$\begin{aligned} A &= (n-p+1)\sigma_n^2(p-q)\left[n\left(\frac{C}{2} - 1\right) - p\frac{C}{4} + q\left(1 - \frac{C}{4}\right) + \frac{C}{2} - 1\right] \\ &\quad + \frac{C}{4}\frac{(n-p+1)^4}{(2n-p-q+2)}\sigma_n^2. \\ B &= 2\frac{(n-p+1)^4}{(p-q)(2n-p-q+2)}\sum_q^{p-1} x_i^2 + (n-p+1)^2 \sum_q^{p-1} x_i^2. \\ C &= 2(p-q)(2n-p-q+2)\sum_p^n x_i^2 - (p-q)(2n-p-q+2)\sum_p^n x_i^2. \end{aligned}$$

  Clearly, $B \geq 0, C \geq 0$.

  Since $p \leq \widetilde{p}$, one has $-p\frac{C}{4} \geq -C\frac{\sqrt{2}}{8}q - \frac{C}{4}\left(1 - \frac{\sqrt{2}}{2}\right)(n+1)$.

  Thus,

$$\begin{aligned} &n\left(\tfrac{C}{2} - 1\right) - p\tfrac{C}{4} + q\left(1 - \tfrac{C}{4}\right) + \tfrac{C}{2} - 1 \geq n\left(C\tfrac{2+\sqrt{2}}{8} - 1\right) + q\left(1 - C\tfrac{2+\sqrt{2}}{8}\right) + C\left[\tfrac{2+\sqrt{2}}{8}\right] - 1 \\ &\geq \left(C\tfrac{2+\sqrt{2}}{8} - 1\right)(n-q+1) \geq 0, \end{aligned}$$

  (recall that $C = 16\log 2 - 8$). As a consequence, $A \geq 0$ and finally $x \leq G_n T_2$.

- if $p > \widetilde{p}$:

$$G_n T - x = G_n T_1 - x =: A + B + C$$

where

$$\begin{aligned}
A &:= \sigma_n^2(p-q)[(n-p+1)^2+(p-q)\frac{(2n-p-q+2)^2}{(n-p+1)} - (n-p+1)(n-q+1)] \\
&= \sigma_n^2\frac{(p-q)^2}{n-p+1}(n-q+1)(3n-2p-q+3) \geq 0 \\
B &:= 2(n-p+1)^2\sum_q^{p-1} x_i^2 + (n-p+1)^2\sum_q^{p-1} x_1^2 \geq 0 \\
C &= [2(p-q)(2n-p-q+2)-(p-q)(2n-p-q+2)]\sum_p^n x_i^2 \geq 0.
\end{aligned}$$

Thus, we find again $x \leq G_n T$.

Finally, we need to check that $x \geq 0$.

Clearly, $x \geq \sigma_n^2(p-q)(n-p+1)(n-q+1)-(n-p+1)^2\sum_q^{p-1} x_i^2$, so that a sufficient condition for $x \geq 0$ is:

$$(p-q)\sigma_n^2 \geq \sum_q^{p-1} x_i^2.$$

Assumption 2) entails that $|x_q| \leq e^{-q(\beta-\varepsilon_q)}$. Since the sequence (ε_n) is non-increasing, we get that $|x_i| \leq e^{-q(\beta-\varepsilon_q)}$ as well for $i \geq q$ and large enough n. Thus :

$$\sum_q^{p-1} x_i^2 \leq (p-q)e^{-2q(\beta-\varepsilon_q)} \leq (p-q)\sigma_n^2,$$

because $e^{-q(\beta-\varepsilon_q)} = \sigma_n^{\frac{\beta-\varepsilon_q}{\beta-1/n}} \leq \sigma_n$ (recall that $\varepsilon_q < \frac{1}{n}$ - see remark 2).

Theorem 5 applies, and we need an estimate of $\frac{x^2}{2G_n}$.

In that view, we shall obtain a finer estimate of $\sum_q^{p-1} x_i^2$ than the one above.

For $j \geq 0$, one has, by assumption 2):

$$\begin{aligned}
|x_{q+j}| &\leq e^{-(q+j)(\beta-\varepsilon_{q+j})} \\
&\leq e^{-q(\beta-\varepsilon_q)}e^{-q(\varepsilon_q-\varepsilon_{q+j})}e^{-j(\beta-\varepsilon_{q+j})}
\end{aligned}$$

Since (ε_n) is non-increasing, $\varepsilon_q-\varepsilon_{q+j} \geq 0$. In addition, since $(\varepsilon_n)\xrightarrow[n\to\infty]{}0$, for sufficiently large n, $\varepsilon_{q+j} < \frac{\beta}{2}$.

Thus:

$$|x_{q+j}| \leq e^{-q(\beta-\varepsilon_q)} e^{-j\frac{\beta}{2}} = \sigma_n e^{-j\frac{\beta}{2}},$$

and:

$$\sum_{j=0}^{p-q-1} x_{q+j}^2 \leq K\sigma_n^2,$$

where K is a positive finite constant. As a consequence:

$$\begin{aligned} x &\geq \sigma_n^2(p-q)(n-p+1)(n-q+1) - K(n-p+1)^2\sigma_n^2 \\ &\geq \sigma^2(n-p+1)[(p-q)(n-q+1) - K(n-p+1)] \\ &\geq \sigma_n^2(n-p+1)(n-q+1)(p-q-K). \end{aligned}$$

Write $G_n =: A + B + C$, with

$$\begin{aligned} A &:= C\sigma_n^4(p-q)(n-p+1)[(n-p+1)^3 + (p-q)(2n-p-q+2)^2]. \\ B &:= 8\sigma_n^2(n-p+1)^4 \sum_q^{p-1} x_i^2. \\ C &:= 8\sigma_n^2(p-q)^2(2n-p-q+2)^2 \sum_p^n x_i^2. \end{aligned}$$

If $p \leq \widetilde{p}$, one has $(p-q)(2n-p-q+2) \leq (n-p+1)^2$, and:

$$\begin{aligned} A &\leq C\sigma_n^4(p-q)(n-p+1)[(n-p+1)^3 + (n-p+1)^2(n-p+1+n-q+1)] \\ &\leq 3C\sigma_n^4(p-q)(n-p+1)^3(n-q+1). \end{aligned}$$

With K' denoting again a positive finite constant that may change from line to line, we have:
$B \leq K\sigma_n^4(n-p+1)^4 \leq K'A$, and
$C \leq 8\sigma_n^2(p-q)^2(2n-p-q+2)^2(n-p+1)\sigma_n^2 \leq K'A$,

Finally, $G_n \leq K'\sigma_n^4(p-q)(n-p+1)^3(n-q+1)$, and:

$$\begin{aligned} \frac{x^2}{2G_n} &\geq \frac{\sigma_n^4(n-p+1)^2(n-q+1)^2(p-q-K)^2}{K'\sigma_n^4(n-p+1)^3(n-q+1)(p-q)} \\ &\geq \frac{1}{K'}\frac{n-q+1}{n-p+1}\frac{(p-q-K)^2}{p-q} \\ &\geq K'\frac{(p-q-K)^2}{p-q}. \end{aligned}$$

Thus, whenever $(p-q) > a\log(n)$ for some $a > 1$, the Borel-Cantelli lemma implies that, for n large enough, $\mathcal{L}(q) < \mathcal{L}(p)$ almost surely.

If $p > \widetilde{p}$,

$$\begin{aligned}
A &\leq 5C\sigma_n^4(p-q)^2(n-p+1)(n-q+1)^2, \\
B &\leq A, \\
C &\leq A,
\end{aligned}$$

and thus

$$\frac{x^2}{2G_n} \geq K'(n-p+1)\frac{(p-q-K)^2}{(p-q)^2}.$$

Since, by assumption on p, $n-p+1 > b\log(n)$ with $b > 1$, the results follows again by the Borel-Cantelli lemma.

Proof of (12)

We want to show that, under assumptions 1) - 4), almost surely, $\mathcal{L}_n(q(n)) - \mathcal{L}_n(p(n)) \leq 0$ for large enough n, for any sequence $p(n)$ such that $1 \leq p(n) < q(n) - \max(a\log(n), \theta_n)$. Take such a sequence $p(n)$. Note that, in particular, $p(n) < q(n)$. Again, we shall write $\mathcal{L}$, p, q in place of $\mathcal{L}_n$, $p(n)$, $q(n)$.

$\mathcal{L}(q) - \mathcal{L}(p) > 0$ is equivalent to:

$$X := (n-q+1)^2\sum_{i=p}^{q-1} y_i^2 + (p-q)(2n-p-q+2)\sum_{i=q}^{n} y_i^2 < 0.$$

Define:

$$\begin{aligned}
X_i &= \sigma_n^2(n-q+1)^2\Omega_i & &\text{for} \quad i = p\ldots q-1, \\
X_i &= \sigma_n^2(p-q)(2n-p-q+2)\Omega_i & &\text{for} \quad i = q\ldots n, \\
X_i &= 0 & &\text{for} \quad i = 0\ldots p-1.
\end{aligned}$$

with Ω_i as above. Set:

$$\begin{aligned}
S_n &:= \sum_{i=1}^{n} X_i \\
&= X - (n-q+1)^2\sum_{i=p}^{q-1} x_i^2 + (q-p)(2n-p-q+2)\sum_{i=q}^{n} x_i^2 \\
&\quad + \sigma_n^2(q-p)(n-p+1)(n-q+1) \\
&=: X - x
\end{aligned}$$

Thus $X < 0$ iff $S_n < -x$.

In view of applying Theorem 5, we need to check that $x \geq 0$. We can easily bound the term $\sum_{i=q}^{n} x_i^2$:

$$\begin{aligned}
\sum_{i=q}^{n} x_i^2 &\leq \sum_{i=q}^{n} e^{-2i(\beta-\varepsilon_i)} \\
&= \sum_{j=0}^{n-q} e^{-2(j+q)(\beta-\varepsilon_{j+q})} \\
&= \sum_{j=0}^{n-q} e^{-2q(\beta-\varepsilon_{j+q})} e^{-2j(\beta-\varepsilon_{j+q})} \\
&= e^{-2q(\beta-\varepsilon_q)} \sum_{j=0}^{n-q} e^{-2q(\varepsilon_q-\varepsilon_{j+q})} e^{-2j(\beta-\varepsilon_{j+q})} \\
&\leq \sigma_n^2 \sum_{j=0}^{n-q} e^{-2j(\beta-\varepsilon_{j+q})} \\
&\leq \sigma_n^2
\end{aligned}$$

where we have used the fact that the sequence (ε_n) is non increasing. Then:

$$x \geq (n-q+1)^2 \sum_{i=p}^{q-1} x_i^2 - \sigma_n^2 (q-p)[2n-p-q+2+(n-p+1)(n-q+1)]$$

Note that:

$$n - q + 1 > n - q \geq n(1-\rho),$$

where ρ is defined in Remark 1. Also

$$\begin{aligned}
2n-p-q+2+(n-p+1)(n-q+1) &\leq 2n+1-\rho' n + n(n-\rho' n+1) \\
&= n^2(1-\rho') + n(3-\rho') + 1 \\
&\leq n^2(1-\hat{\rho}'),
\end{aligned}$$

for any fixed $\hat{\rho}' < \rho'$ provided n is sufficiently large, where ρ' is defined in Remark 1.

As a consequence,

$$x \geq n^2 \left[(1-\rho)^2 \sum_{i=p}^{q-1} x_i^2 - (1-\hat{\rho}')(q-p)\sigma_n^2 \right].$$

Condition 4 then entails:

$$x \geq n^2 (q-p)\sigma_n^2 (1-\rho')(b - \frac{1-\hat{\rho}'}{1-\rho'}).$$

For a fixed $b > 1$, one may choose $\hat{\rho}'$ so that $\eta := (b - \frac{1-\hat{\rho}'}{1-\rho'}) > 0$, and thus

$$x \geq \eta(1-\hat{\rho}')n^2(q-p)\sigma_n^2 > 0. \tag{15}$$

Let us now compute T and G.

- for $i = p \ldots q-1$:

  $\mathbb{E}(e^{tX_i}) \leq e^{g_i \frac{t^2}{2}}$ for $t \in \left[-\frac{1}{4\sigma_n^2(n-q+1)^2}, \frac{1}{4\sigma_n^2(n-q+1)^2} \right] =: [-T_1, T_1]$

  and $g_i = C\sigma_n^4(n-q+1)^4 + 8\sigma_n^2 x_i^2 (n-q+1)^4$.

- for $i = q \ldots n$:

  $\mathbb{E}(e^{tX_i}) \leq e^{g_i \frac{t^2}{2}}$

  for $t \in \left[\frac{1}{4\sigma_n^2(p-q)(2n-p-q+2)}, -\frac{1}{4\sigma_n^2(p-q)(2n-p-q+2)} \right] =: [-T_2, T_2]$

  and $g_i = C\sigma_n^4(q-p)^2(2n-p-q+2)^2 + 8\sigma_n^2 x_i^2 (q-p)^2(2n-p-q+2)^2$.

$$\begin{aligned} G_n := \sum_{i=1}^{n} g_i &= C\sigma_n^4(q-p)(n-q+1)[(n-q+1)^3 + (q-p)(2n-p-q+2)^2] \\ &+ 8\sigma_n^2[(n-q+1)^4 \sum_{p}^{q-1} x_i^2 + (q-p)^2(2n-p-q+2)^2 \sum_{i=q}^{n} x_i^2] \end{aligned}$$

Let us compute the minimum of T_1 and T_2.

$T_2 < T_1$ iff $(q-p)(2n-p-q+2)-(n-q+1)^2 > 0$. Reasoning as above, one finds:

$$\begin{aligned} T &= T_2 \quad \text{if} \quad 1 \le p < \widetilde{p} := \sqrt{2}q - \left(\sqrt{2}-1\right)(n+1) \\ &= T_1 \quad \text{if} \quad \widetilde{p} \le p < q. \end{aligned}$$

Note that the case $T = T_2$ may or may not occur, depending on the value of q: Indeed, $\widetilde{p} \ge 1$ implies that $q \ge (1-\frac{\sqrt{2}}{2})n+1$. For instance, if $\sigma_n = 2^{-n/2}$, this requires $\beta \le \frac{1}{2-\sqrt{2}} \approx 1.7$.

The next step is to compare x and $G_n T$.

- if $p \ge \widetilde{p}$:

$$\begin{aligned} G_n T_1 - x &= \frac{C}{4}\sigma_n^2(q-p)[(n-q+1)^2+(q-p)\frac{(2n-p-q+2)^2}{n-q+1}] \\ &+ 2(n-q+1)^2\sum_p^{q-1} x_i^2 + 2\frac{(q-p)^2(2n-p-q+2)^2}{(n-q+1)^2}\sum_q^n x_i^2 \\ &- (n-q+1)^2\sum_p^{q-1} x_i^2 + (q-p)(2n-p-q+2)\sum_q^n x_i^2 \\ &+ \sigma_n^2(q-p)(n-p+1)(n-q+1) \\ &\ge (n-q+1)^2\sum_p^{q-1} x_i^2 \\ &\ge 0. \end{aligned}$$

- when $p < \widetilde{p}$, the comparison between x and $G_n T$ is more complex, and requires to distinguish further subcases. In view of applying theorem 5, it is simpler to estimate directly both $\frac{x^2}{2G_n}$ and $\frac{T_2 x}{2}$.

At this point, it thus remains to show that $\frac{x^2}{2G_n}$ tends "sufficiently fast" to infinity (which is needed in both cases $p < \widetilde{p}$ and $p \ge \widetilde{p}$), and that the same is true for $\frac{T_2 x}{2}$ (this is needed only $p < \widetilde{p}$).

We begin with $\frac{T_2 x}{2}$. In view of (15):

$$\begin{aligned} \frac{T_2 x}{2} &\ge \frac{\eta(1-\hat{\rho}')n^2(q-p)\sigma_n^2}{8\sigma_n^2(q-p)(2n-p-q+2)} \\ &\ge \frac{\eta(1-\hat{\rho}')n}{16}, \end{aligned}$$

which is sufficient for the Borel-Cantelli lemma to apply.

Let us now consider $\frac{x^2}{2G_n}$. We first compute a crude upper bound to G_n. Using that $\sum_{i=q}^{n} x_i^2 \leq \sigma_n^2$, one gets:

$$G_n \leq K\left(n^4(q-p)\sigma_n^4 + n^4\sigma_n^2\sum_{i=p}^{q-1} x_i^2 + n^3(q-p)\sigma_n^4\right).$$

Using again (15):

$$G_n \leq Kn^4\sigma_n^2\sum_{i=p}^{q-1} x_i^2.$$

As a consequence:

$$\frac{x^2}{2G_n} \geq \frac{\left[(1-\rho)^2\sum_{i=p}^{q-1} x_i^2 - (1-\hat{\rho}')(q-p)\sigma_n^2\right]^2}{K\sigma_n^2\sum_{i=p}^{q-1} x_i^2}.$$

The function $F_z(y) = \frac{(z-y)^2}{zy}$ is decreasing. Thus:

$$\begin{aligned}
\frac{\left[(1-\rho)^2\sum_{i=p}^{q-1} x_i^2 - (1-\hat{\rho}')(q-p)\sigma_n^2\right]^2}{K\sigma_n^2\sum_{i=p}^{q-1} x_i^2} &= (q-p)\frac{(1-\rho)^2(1-\hat{\rho}')}{K}F_{(1-\rho)^2\sum_{i=p}^{q-1} x_i^2}\left((1-\hat{\rho}')(q-p)\sigma_n^2\right) \\
&\geq (q-p)\frac{(1-\rho)^2(1-\hat{\rho}')}{K}F_{(1-\rho)^2\sum_{i=p}^{q-1} x_i^2}\left(\frac{(1-\hat{\rho}')(1-\rho)^2}{b(1-\rho')}\sum_{i=p}^{q-1} x_i^2\right) \\
&= (q-p)\frac{(1-\rho)^2(1-\hat{\rho}')}{K}b(1-\rho')\frac{(1-\frac{1-\hat{\rho}'}{b(1-\rho')})^2}{1-\hat{\rho}'}.
\end{aligned}$$

Recall that $(1-\frac{1-\hat{\rho}'}{b(1-\rho')}) > 0$. We finally get:

$$\frac{x^2}{2G_n} \geq K(q-p).$$

Since $q-p > a\log(n)$, the result follows by the Borel-Cantelli lemma. $\square$

References

[1] R. Devore and B. Lucier, "Fast wavelet techniques for near optimal image processing," *IEEE Military Comm. Conf.*, vol. 2–12, 1992.

[2] D. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, 1995.

[3] D. Kerkyacharian, D. Picard, and K. Tribouley, "l_p adaptative density estimation," *Bernouilli*, vol. 2, pp. 229–247, 1996.

[4] P. Hall, G. Kerkyacharian, and D. Picard, "Block threshold rules for curve estimation using kernel and wavelet methods," *Ann. Statist.*, vol. 26, no. 3, pp. 922–942, 1998.

[5] A. Antoniadis, J. Bigot, and T. Sapatinas, "Wavelet estimators in non-parametric regression: a comparative simulation study," *J. Statist. Software 6*, vol. 6, pp. 1–83, 2001.

[6] R. Coifman and D. Donoho, *Translation invariant de-noising.* SpringerVerlag, 1995, pp. 125–150.

[7] B. Vidakovic, "Nonlinear wavelet shrinkage with Bayes rules and Bayes factors," *Ann. Statist.*, vol. 93, pp. 173–179, 1998.

[8] ——, *Statistical Modeling by Wavelets.* New York: Wiley- Interscience, 1999.

[9] S. Jaffard, "Wavelet techniques in multifractal analysis," in *Wavelet and multifractal analysis, Institut d'Etudes Scientifiques de Cargse*, 2004.

[10] R. R. Coifman and A. Sowa, "Combining variational calculus and wavelets for image enhancement," *Appl. Comput. Harmon. Anal.*, vol. 9, pp. 1–18, 2000.

[11] S. Durand and J. Froment, "Reconstruction of wavelet coefficients using total variation minimization," *SIAM Journal of Scientific Computing*, vol. 24, no. 5, pp. 1754–1767, 2003.

[12] K. Falconer, "The local structure of random processes," *J. London Math. Soc.*, vol. 2, no. 67, pp. 657–672, 2003.

[13] S. Stoev and M. Taqqu, "Stochastic properties of the linear multifractional stable motion," *Adv. Appl. Probab.*, vol. 36, pp. 1085–1115, 2004.

[14] K. Falconer and J. Lévy Véhel, "Multifractional, multistable, and other processes with prescribed local form," *J. Theoret. Probab*, vol. 22, pp. 375–401, 2009.

[15] T. Hsung, D. Lun, and W. Siu, "Denoising by singularity detection," *IEEE Trans. Signal Process.*, vol. 47, no. 11, pp. 3139–3144, 1999.

[16] F. Luisier, T. Blu, and M. Unser, "A new sure approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Trans. Signal Process.*, vol. 16, no. 3, pp. 593–606, 2007.

[17] D. Picard and K. Tribouley, "Adaptative confidence interval for pointwise curve estimation," *Ann. Stat.*, vol. 28, no. 1, pp. 298–335, 2000.

[18] S. Mallat, *A wavelet tour of signal processing.* Academic Press, 1998.

[19] G. Strang and T. Nguyen, *Wavelets and filter banks.* Wellesley-Cambridge Press, 1996.

[20] O. M. Nielsen, "Wavelets in scientific computing," Ph.D. dissertation, Department of Mathematical Modelling Technical university of Denmark, 1998. [Online]. Available: http://www2.imm.dtu.dk/documents/ftp/phdliste/phd43.abstract.html

[21] I. N. Guelfand and N. Y. Vilenkin, *Generalized Functions.* Academic Press, 1964, vol. 4. Applications of Harmonic Analysis.

[22] M. R. Leadbetter, G. Lindgren, and H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes.* Springer, 1983.

[23] V. Petrov, *Limit Theorems of Probability Theory.* Oxford Science Publication, 1995.

[24] S. Bianchi, "Pathwise identification of the memory function of multifractional Brownian motion with application to finance," *Int. J. Theor. and Appl. Finance*, vol. 2, no. 8, pp. 255–281, 2005.

[25] S. Gaci, J. Lévy Véhel, C. Keylock, , J. Wanliss, and D. Schertzer (Eds.), "Special issue on multifractional Brownian motions in geosciences," *Nonlin. Processes Geophys.*, 2013. [Online]. Available: http://www.nonlin-processes-geophys.net/special_issue142.html

[26] M. Li, W. Zhao, and S. Chen, "mBm-based scalings of traffic propagated in Internet," *Mathematical Problems in Engineering*, vol. 93, 2011, article ID 389803, doi:10.1155/2011/389803.

[27] N. Betrouni, N. Makni, A. Iancu, P. Puech, and S. Mordon, "Computer assisted treatment of the prostate," *Int. J. Computer Assisted Radiology Surgery*, vol. 7, no. 1, pp. 13–18, 2012.